

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problems Mailbox.**

THIS PAGE BLANK (USPTO)

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner #: _____ Date: _____
 Art Unit: _____ Phone Number 30 _____ Serial Number: _____
 Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

STAFF USE ONLY

	Type of Search	Vendors and cost where applicable
Searcher: _____	NA Sequence (#) _____	STN _____
Searcher Phone #: If Contact: Sheppard	AA Sequence (#) _____	Dialog _____
Searcher Location: tel: 308-4499	Structure (#) _____	Questel/Orbi: _____
Date Searcher Picked Up: _____	Bibliographic _____	Dr.Link _____
Date Completed: 8/11/01	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: _____	Fulltext _____	Sequence Systems _____
Clerical Prep Time: _____	Patent Family _____	WWW/Internet _____
Online Time: _____	Other _____	Other (specify) _____

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 01:25:08 ; Search time 6114.61 Seconds
(without alignments)
3149.091 Million cell updates/sec

Title: US-09-402-713A-1
Perfect score: 2037
Sequence: 1 aqaagctgcatcagaaaaa.....caataaagaatttacaaga 2037

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues 20456230
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST.*

- 1: gb_estl1.*
- 2: gb_est2.*
- 3: gb_est3.*
- 4: gb_est4.*
- 5: gb_est5.*
- 6: gb_est6.*
- 7: gb_est7.*
- 8: gb_est8.*
- 9: gb_est9.*
- 10: gb_est10.*
- 11: gb_est11.*
- 12: gb_est12.*
- 13: gb_est13.*
- 14: gb_est14.*
- 15: gb_est15.*
- 16: gb_est16.*
- 17: gb_est17.*
- 18: gb_est18.*
- 19: gb_est19.*
- 20: gb_est20.*
- 21: gb_est21.*
- 22: gb_est22.*
- 23: gb_est23.*
- 24: gb_est24.*
- 25: gb_est25.*
- 26: gb_est26.*
- 27: gb_est27.*
- 28: gb_est28.*
- 29: gb_est29.*
- 30: gb_est30.*
- 31: gb_est31.*
- 32: gb_est32.*
- 33: gb_est33.*
- 34: gb_est34.*
- 35: gb_est35.*
- 36: gb_est36.*
- 37: gb_est37.*
- 38: gb_est38.*
- 39: gb_est39.*
- 40: gb_est40.*
- 41: gb_est41.*
- 42: gb_est42.*
- 43: gb_est43.*
- 44: gb_est44.*
- 45: gb_est45.*
- 46: gb_est46.*
- 47: gb_est47.*
- 48: gb_est48.*
- 49: gb_est49.*
- 50: gb_est50.*
- 51: gb_est51.*
- 52: gb_est52.*
- 53: gb_est53.*
- 54: gb_est54.*
- 55: gb_est55.*
- 56: gb_est56.*
- 57: gb_est57.*
- 58: gb_est58.*
- 59: gb_est59.*
- 60: gb_est60.*
- 61: gb_est61.*
- 62: gb_est62.*
- 63: gb_est63.*
- 64: gb_est64.*
- 65: gb_est65.*
- 66: gb_est66.*
- 67: gb_est67.*
- 68: gb_est68.*
- 69: gb_est69.*
- 70: gb_est70.*
- 71: gb_est71.*
- 72: gb_est72.*
- 73: gb_est73.*
- 74: gb_est74.*
- 75: gb_est75.*
- 76: gb_est76.*
- 77: gb_est77.*
- 78: gb_est78.*
- 79: gb_est79.*
- 80: gb_est80.*
- 81: gb_est81.*
- 82: gb_est82.*
- 83: gb_est83.*
- 84: gb_est84.*
- 85: gb_est85.*
- 86: gb_est86.*
- 87: gb_est87.*
- 88: gb_est88.*
- 89: gb_est89.*
- 90: gb_est90.*
- 91: gb_est91.*
- 92: gb_est92.*
- 93: gb_est93.*
- 94: gb_est94.*
- 95: gb_est95.*
- 96: gb_est96.*
- 97: gb_est97.*
- 98: gb_est98.*
- 99: gb_est99.*
- 100: gb_est100.*
- 101: gb_est101.*
- 102: gb_est102.*
- 103: gb_est103.*
- 104: gb_est104.*
- 105: gb_est105.*
- 106: gb_est106.*
- 107: gb_est107.*
- 108: gb_est108.*
- 109: gb_est109.*
- 110: gb_est110.*
- 111: gb_est111.*
- 112: gb_est112.*
- 113: gb_est113.*
- 114: gb_est114.*
- 115: gb_est115.*
- 116: gb_est116.*
- 117: gb_est117.*
- 118: gb_est118.*
- 119: gb_est119.*
- 120: gb_est120.*
- 121: gb_est121.*
- 122: gb_est122.*
- 123: gb_est123.*
- 124: gb_est124.*
- 125: gb_est125.*
- 126: gb_est126.*
- 127: gb_est127.*
- 128: gb_est128.*
- 129: gb_est129.*
- 130: gb_est130.*
- 131: gb_est131.*
- 132: gb_est132.*
- 133: gb_est133.*
- 134: gb_est134.*
- 135: gb_est135.*
- 136: gb_est136.*
- 137: gb_est137.*
- 138: gb_est138.*
- 139: gb_est139.*
- 140: gb_est140.*
- 141: gb_est141.*
- 142: gb_est142.*
- 143: gb_est143.*
- 144: gb_est144.*
- 145: gb_est145.*
- 146: gb_est146.*
- 147: gb_est147.*
- 148: gb_est148.*
- 149: gb_est149.*
- 150: gb_est150.*
- 151: gb_est151.*
- 152: gb_est152.*
- 153: gb_est153.*
- 154: gb_est154.*
- 155: gb_est155.*
- 156: gb_est156.*
- 157: gb_est157.*
- 158: gb_est158.*
- 159: gb_est159.*
- 160: gb_est160.*
- 161: gb_est161.*
- 162: gb_est162.*
- 163: gb_est163.*
- 164: gb_est164.*
- 165: gb_est165.*
- 166: gb_est166.*
- 167: gb_est167.*
- 168: gb_est168.*
- 169: gb_est169.*
- 170: gb_est170.*
- 171: gb_est171.*
- 172: gb_est172.*
- 173: gb_est173.*
- 174: gb_est174.*
- 175: gb_est175.*
- 176: gb_est176.*
- 177: gb_est177.*
- 178: gb_est178.*
- 179: gb_est179.*
- 180: gb_est180.*
- 181: gb_est181.*
- 182: gb_est182.*
- 183: gb_est183.*
- 184: gb_est184.*
- 185: gb_est185.*
- 186: gb_est186.*
- 187: gb_est187.*
- 188: gb_est188.*
- 189: gb_est189.*
- 190: gb_est190.*
- 191: gb_est191.*
- 192: gb_est192.*
- 193: gb_est193.*
- 194: gb_est194.*
- 195: gb_est195.*
- 196: gb_est196.*
- 197: gb_est197.*
- 198: gb_est198.*
- 199: gb_est199.*
- 200: gb_est200.*
- 201: gb_est201.*
- 202: gb_est202.*
- 203: gb_est203.*
- 204: gb_est204.*
- 205: gb_est205.*
- 206: gb_est206.*
- 207: gb_est207.*
- 208: gb_est208.*
- 209: gb_est209.*
- 210: gb_est210.*
- 211: gb_est211.*
- 212: gb_est212.*
- 213: gb_est213.*
- 214: gb_est214.*
- 215: gb_est215.*
- 216: gb_est216.*
- 217: gb_est217.*
- 218: gb_est218.*
- 219: gb_est219.*
- 220: gb_est220.*
- 221: gb_est221.*
- 222: gb_est222.*
- 223: gb_est223.*
- 224: gb_est224.*
- 225: gb_est225.*
- 226: gb_est226.*
- 227: gb_est227.*
- 228: gb_est228.*
- 229: gb_est229.*
- 230: gb_est230.*
- 231: gb_est231.*
- 232: gb_est232.*
- 233: gb_est233.*
- 234: gb_est234.*
- 235: gb_est235.*
- 236: gb_est236.*
- 237: gb_est237.*
- 238: gb_est238.*
- 239: gb_est239.*
- 240: gb_est240.*
- 241: gb_est241.*
- 242: gb_est242.*
- 243: gb_est243.*
- 244: gb_est244.*
- 245: gb_est245.*
- 246: gb_est246.*
- 247: gb_est247.*
- 248: gb_est248.*
- 249: gb_est249.*
- 250: gb_est250.*
- 251: gb_est251.*
- 252: gb_est252.*
- 253: gb_est253.*
- 254: gb_est254.*
- 255: gb_est255.*
- 256: gb_est256.*
- 257: gb_est257.*
- 258: gb_est258.*
- 259: gb_est259.*
- 260: gb_est260.*
- 261: gb_est261.*
- 262: gb_est262.*
- 263: gb_est263.*
- 264: gb_est264.*
- 265: gb_est265.*
- 266: gb_est266.*
- 267: gb_est267.*
- 268: gb_est268.*
- 269: gb_est269.*
- 270: gb_est270.*
- 271: gb_est271.*
- 272: gb_est272.*
- 273: gb_est273.*
- 274: gb_est274.*
- 275: gb_est275.*
- 276: gb_est276.*
- 277: gb_est277.*
- 278: gb_est278.*
- 279: gb_est279.*
- 280: gb_est280.*
- 281: gb_est281.*
- 282: gb_est282.*
- 283: gb_est283.*
- 284: gb_est284.*
- 285: gb_est285.*
- 286: gb_est286.*
- 287: gb_est287.*
- 288: gb_est288.*
- 289: gb_est289.*
- 290: gb_est290.*
- 291: gb_est291.*
- 292: gb_est292.*
- 293: gb_est293.*
- 294: gb_est294.*
- 295: gb_est295.*
- 296: gb_est296.*
- 297: gb_est297.*
- 298: gb_est298.*
- 299: gb_est299.*
- 300: gb_est300.*
- 301: gb_est301.*
- 302: gb_est302.*
- 303: gb_est303.*
- 304: gb_est304.*
- 305: gb_est305.*
- 306: gb_est306.*
- 307: gb_est307.*
- 308: gb_est308.*
- 309: gb_est309.*
- 310: gb_est310.*
- 311: gb_est311.*
- 312: gb_est312.*
- 313: gb_est313.*
- 314: gb_est314.*
- 315: gb_est315.*
- 316: gb_est316.*
- 317: gb_est317.*
- 318: gb_est318.*
- 319: gb_est319.*
- 320: gb_est320.*
- 321: gb_est321.*
- 322: gb_est322.*
- 323: gb_est323.*
- 324: gb_est324.*
- 325: gb_est325.*
- 326: gb_est326.*
- 327: gb_est327.*
- 328: gb_est328.*
- 329: gb_est329.*
- 330: gb_est330.*
- 331: gb_est331.*
- 332: gb_est332.*
- 333: gb_est333.*
- 334: gb_est334.*
- 335: gb_est335.*
- 336: gb_est336.*
- 337: gb_est337.*
- 338: gb_est338.*
- 339: gb_est339.*
- 340: gb_est340.*
- 341: gb_est341.*
- 342: gb_est342.*
- 343: gb_est343.*
- 344: gb_est344.*
- 345: gb_est345.*
- 346: gb_est346.*
- 347: gb_est347.*
- 348: gb_est348.*
- 349: gb_est349.*
- 350: gb_est350.*
- 351: gb_est351.*
- 352: gb_est352.*
- 353: gb_est353.*
- 354: gb_est354.*
- 355: gb_est355.*
- 356: gb_est356.*
- 357: gb_est357.*
- 358: gb_est358.*
- 359: gb_est359.*
- 360: gb_est360.*
- 361: gb_est361.*
- 362: gb_est362.*
- 363: gb_est363.*
- 364: gb_est364.*
- 365: gb_est365.*
- 366: gb_est366.*
- 367: gb_est367.*
- 368: gb_est368.*
- 369: gb_est369.*
- 370: gb_est370.*
- 371: gb_est371.*
- 372: gb_est372.*
- 373: gb_est373.*
- 374: gb_est374.*
- 375: gb_est375.*
- 376: gb_est376.*
- 377: gb_est377.*
- 378: gb_est378.*
- 379: gb_est379.*
- 380: gb_est380.*
- 381: gb_est381.*
- 382: gb_est382.*
- 383: gb_est383.*
- 384: gb_est384.*
- 385: gb_est385.*
- 386: gb_est386.*
- 387: gb_est387.*
- 388: gb_est388.*
- 389: gb_est389.*
- 390: gb_est390.*
- 391: gb_est391.*
- 392: gb_est392.*
- 393: gb_est393.*
- 394: gb_est394.*
- 395: gb_est395.*
- 396: gb_est396.*
- 397: gb_est397.*
- 398: gb_est398.*
- 399: gb_est399.*
- 400: gb_est400.*
- 401: gb_est401.*
- 402: gb_est402.*
- 403: gb_est403.*
- 404: gb_est404.*
- 405: gb_est405.*
- 406: gb_est406.*
- 407: gb_est407.*
- 408: gb_est408.*
- 409: gb_est409.*
- 410: gb_est410.*
- 411: gb_est411.*
- 412: gb_est412.*
- 413: gb_est413.*
- 414: gb_est414.*
- 415: gb_est415.*
- 416: gb_est416.*
- 417: gb_est417.*
- 418: gb_est418.*
- 419: gb_est419.*
- 420: gb_est420.*
- 421: gb_est421.*
- 422: gb_est422.*
- 423: gb_est423.*
- 424: gb_est424.*
- 425: gb_est425.*
- 426: gb_est426.*
- 427: gb_est427.*
- 428: gb_est428.*
- 429: gb_est429.*
- 430: gb_est430.*
- 431: gb_est431.*
- 432: gb_est432.*
- 433: gb_est433.*
- 434: gb_est434.*
- 435: gb_est435.*
- 436: gb_est436.*
- 437: gb_est437.*
- 438: gb_est438.*
- 439: gb_est439.*
- 440: gb_est440.*
- 441: gb_est441.*
- 442: gb_est442.*
- 443: gb_est443.*
- 444: gb_est444.*
- 445: gb_est445.*
- 446: gb_est446.*
- 447: gb_est447.*
- 448: gb_est448.*
- 449: gb_est449.*
- 450: gb_est450.*
- 451: gb_est451.*
- 452: gb_est452.*
- 453: gb_est453.*
- 454: gb_est454.*
- 455: gb_est455.*
- 456: gb_est456.*
- 457: gb_est457.*
- 458: gb_est458.*
- 459: gb_est459.*
- 460: gb_est460.*
- 461: gb_est461.*
- 462: gb_est462.*
- 463: gb_est463.*
- 464: gb_est464.*
- 465: gb_est465.*
- 466: gb_est466.*
- 467: gb_est467.*
- 468: gb_est468.*
- 469: gb_est469.*
- 470: gb_est470.*
- 471: gb_est471.*
- 472: gb_est472.*
- 473: gb_est473.*
- 474: gb_est474.*
- 475: gb_est475.*
- 476: gb_est476.*
- 477: gb_est477.*
- 478: gb_est478.*
- 479: gb_est479.*
- 480: gb_est480.*
- 481: gb_est481.*
- 482: gb_est482.*
- 483: gb_est483.*
- 484: gb_est484.*
- 485: gb_est485.*
- 486: gb_est486.*
- 487: gb_est487.*
- 488: gb_est488.*
- 489: gb_est489.*
- 490: gb_est490.*
- 491: gb_est491.*
- 492: gb_est492.*
- 493: gb_est493.*
- 494: gb_est494.*
- 495: gb_est495.*
- 496: gb_est496.*
- 497: gb_est497.*
- 498: gb_est498.*
- 499: gb_est499.*
- 500: gb_est500.*
- 501: gb_est501.*
- 502: gb_est502.*
- 503: gb_est503.*
- 504: gb_est504.*
- 505: gb_est505.*
- 506: gb_est506.*
- 507: gb_est507.*
- 508: gb_est508.*
- 509: gb_est509.*
- 510: gb_est510.*
- 511: gb_est511.*
- 512: gb_est512.*
- 513: gb_est513.*
- 514: gb_est514.*
- 515: gb_est515.*
- 516: gb_est516.*
- 517: gb_est517.*
- 518: gb_est518.*
- 519: gb_est519.*
- 520: gb_est520.*
- 521: gb_est521.*
- 522: gb_est522.*
- 523: gb_est523.*
- 524: gb_est524.*
- 525: gb_est525.*
- 526: gb_est526.*
- 527: gb_est527.*
- 528: gb_est528.*
- 529: gb_est529.*
- 530: gb_est530.*
- 531: gb_est531.*
- 532: gb_est532.*
- 533: gb_est533.*
- 534: gb_est534.*
- 535: gb_est535.*
- 536: gb_est536.*
- 537: gb_est537.*
- 538: gb_est538.*
- 539: gb_est539.*
- 540: gb_est540.*
- 541: gb_est541.*
- 542: gb_est542.*
- 543: gb_est543.*
- 544: gb_est544.*
- 545: gb_est545.*
- 546: gb_est546.*
- 547: gb_est547.*
- 548: gb_est548.*
- 549: gb_est549.*
- 550: gb_est550.*
- 551: gb_est551.*
- 552: gb_est552.*
- 553: gb_est553.*
- 554: gb_est554.*
- 555: gb_est555.*
- 556: gb_est556.*
- 557: gb_est557.*
- 558: gb_est558.*
- 559: gb_est559.*
- 560: gb_est560.*
- 561: gb_est561.*
- 562: gb_est562.*
- 563: gb_est563.*
- 564: gb_est564.*
- 565: gb_est565.*
- 566: gb_est566.*
- 567: gb_est567.*
- 568: gb_est568.*
- 569: gb_est569.*
- 570: gb_est570.*
- 571: gb_est571.*
- 572: gb_est572.*
- 573: gb_est573.*
- 574: gb_est574.*
- 575: gb_est575.*
- 576: gb_est576.*
- 577: gb_est577.*
- 578: gb_est578.*
- 579: gb_est579.*
- 580: gb_est580.*
- 581: gb_est581.*
- 582: gb_est582.*
- 583: gb_est583.*
- 584: gb_est584.*
- 585: gb_est585.*
- 586: gb_est586.*
- 587: gb_est587.*
- 588: gb_est588.*
- 589: gb_est589.*
- 590: gb_est590.*
- 591: gb_est591.*
- 592: gb_est592.*
- 593: gb_est593.*
- 594: gb_est594.*
- 595: gb_est595.*
- 596: gb_est596.*
- 597: gb_est597.*
- 598: gb_est598.*
- 599: gb_est599.*
- 600: gb_est600.*
- 601: gb_est601.*
- 602: gb_est602.*
- 603: gb_est603.*
- 604: gb_est604.*
- 605: gb_est605.*
- 606: gb_est606.*
- 607: gb_est607.*
- 608: gb_est608.*
- 609: gb_est609.*
- 610: gb_est610.*
- 611: gb_est611.*
- 612: gb_est612.*
- 613: gb_est613.*
- 614: gb_est614.*
- 615: gb_est615.*
- 616: gb_est616.*
- 617: gb_est617.*
- 618: gb_est618.*
- 619: gb_est619.*
- 620: gb_est620.*
- 621: gb_est621.*
- 622: gb_est622.*
- 623: gb_est623.*
- 624: gb_est624.*
- 625: gb_est625.*
- 626: gb_est626.*
- 627: gb_est627.*
- 628: gb_est628.*
- 629: gb_est629.*
- 630: gb_est630.*
- 631: gb_est631.*
- 632: gb_est632.*
- 633: gb_est633.*
- 634: gb_est634.*
- 635: gb_est635.*
- 636: gb_est636.*
- 637: gb_est637.*
- 638: gb_est638.*
- 639: gb_est639.*
- 640: gb_est640.*
- 641: gb_est641.*
- 642: gb_est642.*
- 643: gb_est643.*
- 644: gb_est644.*
- 645: gb_est645.*
- 646: gb_est646.*
- 647: gb_est647.*
- 648: gb_est648.*
- 649: gb_est649.*
- 650: gb_est650.*
- 651: gb_est651.*
- 652: gb_est652.*
- 653: gb_est653.*
- 654: gb_est654.*
- 655: gb_est655.*
- 656: gb_est656.*
- 657: gb_est657.*
- 658: gb_est658.*
- 659: gb_est659.*
- 660: gb_est660.*
- 661: gb_est661.*
- 662: gb_est662.*
- 663: gb_est663.*
- 664: gb_est664.*
- 665: gb_est665.*
- 666: gb_est666.*
- 667: gb_est667.*
- 668: gb_est668.*
- 669: gb_est669.*
- 670: gb_est670.*
- 671: gb_est671.*
- 672: gb_est672.*
- 673: gb_est673.*
- 674: gb_est674.*
- 675: gb_est675.*
- 676: gb_est676.*
- 677: gb_est677.*
- 678: gb_est678.*
- 679: gb_est679.*
- 680: gb_est680.*
- 681: gb_est681.*
- 682: gb_est682.*
- 683: gb_est683.*
- 684: gb_est684.*
- 685: gb_est685.*
- 686: gb_est686.*
- 687: gb_est687.*
- 688: gb_est688.*
- 689: gb_est689.*
- 690: gb_est690.*
- 691: gb_est691.*
- 692: gb_est692.*
- 693: gb_est693.*
- 694: gb_est694.*
- 695: gb_est695.*
- 696: gb_est696.*
- 697: gb_est697.*
- 698: gb_est698.*
- 699: gb_est699.*
- 700: gb_est700.*
- 701: gb_est701.*
- 702: gb_est702.*
- 703: gb_est703.*
- 704: gb_est704.*
- 705: gb_est705.*
- 706: gb_est706.*
- 707: gb_est707.*
- 708: gb_est708.*
- 709: gb_est709.*
- 710: gb_est710.*
- 711: gb_est711.*
- 712: gb_est712.*
- 713: gb_est713.*
- 714: gb_est714.*
- 715: gb_est715.*
- 716: gb_est716.*
- 717: gb_est717.*
- 718: gb_est718.*
- 719: gb_est719.*
- 720: gb_est720.*
- 721: gb_est721.*
- 722: gb_est722.*
- 723: gb_est723.*
- 724: gb_est724.*
- 725: gb_est725.*
- 726: gb_est726.*
- 727: gb_est727.*
- 728: gb_est728.*
- 729: gb_est729.*
- 730: gb_est730.*
- 731: gb_est731.*
- 732: gb_est732.*
- 733: gb_est733.*
- 734: gb_est734.*
- 735: gb_est735.*
- 736: gb_est736.*
- 737: gb_est737.*
- 738: gb_est738.*
- 739: gb_est739.*
- 740: gb_est740.*
- 741: gb_est741.*
- 742: gb_est742.*
- 743: gb_est743.*
- 744: gb_est744.*
- 745: gb_est745.*
- 746: gb_est746.*
- 747: gb_est747.*
- 748: gb_est748.*
- 749: gb_est749.*
- 750: gb_est750.*
- 751: gb_est751.*
- 752: gb_est752.*
- 753: gb_est753.*
- 754: gb_est754.*
- 755: gb_est755.*
- 756: gb_est756.*
- 757: gb_est757.*
- 758: gb_est758.*
- 759: gb_est759.*
- 760: gb_est760.*
- 761: gb_est761.*
- 762: gb_est762.*
- 763: gb_est763.*
- 764: gb_est764.*
- 765: gb_est765.*
- 766: gb_est766.*
- 767: gb_est767.*
- 768: gb_est768.*
- 769: gb_est769.*
- 770: gb_est770.*
- 771: gb_est771.*
- 772: gb_est772.*
- 773: gb_est773.*
- 774: gb_est774.*
- 775: gb_est775.*
- 776: gb_est776.*
- 777: gb_est777.*
- 778: gb_est778.*
- 779: gb_est779.*
- 780: gb_est780.*
- 781: gb_est781.*
- 782: gb_est782.*
- 783: gb_est783.*
- 784: gb_est784.*
- 785: gb_est785.*
- 786: gb_est786.*
- 787: gb_est787.*
- 788: gb_est788.*
- 789: gb_est789.*
- 790: gb_est790.*
- 791: gb_est79

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
c 1	613.8	30.1	654	227	AQ319247	RPC111-10
c 2	512.2	25.1	856	21	A1557225	PT2.1.15
c 3	399.6	19.6	441	170	BF858890	RC5-FT019
c 4	394	19.3	402	9	AA578773	nh24a04.s
c 5	362	17.8	394	225	AQ206972	HS-3238_B
c 6	306.2	15.0	332	170	BF858286	RC5-FT019
c 7	260.4	12.8	290	147	BF373619	MR0-FT017
c 8	226.8	11.1	282	147	BF373581	MR0-FT017
c 9	214.4	10.5	226	170	BF858371	RC5-FT019
c 10	167	8.2	167	147	BF373406	IL2-FT015
c 11	138.2	6.8	451	226	AQ246715	HS-2059_B
c 12	86.4	4.2	657	21	A1557495	PT2.1.17_H
c 13	63	3.1	1101	219	CNS0039G	
c 14	56	2.7	1001	219	CNS01400	
c 15	52.6	2.6	1101	219	CNS0182P	
c 16	52.2	2.6	1101	219	CNS0100X	
c 17	51.6	2.5	1101	219	CNS0000D1	
c 18	49.2	2.4	1204	219	CNS016E2	
c 19	48.2	2.4	1101	219	CNS017KX	
c 20	47.8	2.3	1101	219	CNS0039G	
c 21	47.6	2.3	529	225	AQ173559	HS-3202_A
c 22	47.6	2.3	1101	219	CNS0039E	
c 23	47.6	2.3	1101	219	CNS00EVL	
c 24	47.6	2.3	1101	219	CNS0106X	
c 25	47.2	2.3	1201	219	CNS0160R	
c 26	46.8	2.3	654	223	AQ046642	RPC111-35
c 27	45.2	2.2	1086	219	CNS00YXX	
c 28	44.8	2.2	938	219	CNS006TJ	
c 29	44.8	2.2	943	219	CNS002IM	
c 30	44.8	2.2	1125	106	AL547503	
c 31	44.4	2.2	926	219	CNS008LM	
c 32	44.4	2.2	1101	219	CNS017KT	
c 33	43.6	2.1	1101	219	CNS00LT2	
c 34	43.2	2.1	352	12	AA791841	vs65bll.r
c 35	43.2	2.1	656	119	AW691330	NF043E12S
c 36	43.2	2.1	997	219	CNS005TE	
c 37	43.2	2.1	1101	219	CNS017KE	
c 38	43	2.1	928	219	CNS0076D	
c 39	43	2.1	930	219	CNS00D23	
c 40	42.8	2.1	821	142	BE960968	601648521
c 41	42.8	2.1	1101	219	CNS0039L	
c 42	42.6	2.1	859	219	CNS004YY	
c 43	42.6	2.1	1077	222	CNS05AA9	
c 44	42.4	2.1	734	219	CNS010MP	
c 45	42.4	2.1	1092	220	CNS020K7	

ALIGNMENTS

RESULT	1
LOCUS	AQ319247/c
DEFINITION	AQ319247 654 bp DNA GSS 06-MAY-1999
ACCESSION	RPC111-108L4.TV RPC1-11 Homo sapiens genomic clone RPC1-11-108L4, DNA sequence.
VERSION	AQ319247.1 GI:4052212
KEYWORDS	GSS.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 654)
TITLE	Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P., and Venter, J.C.
JOURNAL	Use of human BAC End Sequences for Sequence-Ready Map Building Unpublished (1998)

COMMENT

Contact: Shaying Zhao, William Nierman, Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: hbe@tigr.org
 Clones are derived from the human BAC library RPC1-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html
 Seq primer: 17
 Class: BAC ends.

FEATURES

Location/Qualifiers
 1..654
 /organism="Homo sapiens"
 /db_xref="GDB:754135"
 /db_xref="taxon:9606"
 /clone="RPC1-11-108L4"
 /clone_lib="RPC1-11"
 /sex="Male"
 /cell_type="Lymphocytes"
 /note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
 RPC111 Human Male BAC Library"

BASE COUNT 213 a 100 c 139 g 202 t
 ORIGIN

Query Match 30.1%; Score 613.8; DB 227; Length 654;
 Best Local Similarity 98.5%; Pred. No. 1.2e-153;
 Matches 645; Conservative 4; Mismatches 3; Indels 3; Gaps 3;

QY 1195 tgccattctcaagacctcaaatgtcattccatttaataatcacagattacattttttt 1254
 DB 654 TGCCCATTCCTCAAGACCTCAAAATGTCATTCCATTAAATATCACAGGATTAACTTTT 595
 QY 1255 ttaacctgagaagaattcaatgttacatgcagctatgggaatttaataattatgttt 1314
 DB 594 TTAACCTGGGAAGAAATTCAAATGTTACATGACGCTATGGGAATTAATATCATATTTT 535
 QY 1315 tccagtgcagaagatgactgaagtcctttatccctccctttgtttgtttttttccagta 1374
 DB 534 TCCAGTGCAGAAATGACTAAGTCCCTTTATCCCTCCCTTTGTTGATTTTTCAGTA 475
 QY 1375 taaagttaaaatgcttagccttgtagctgaggtgtatatacag-cacagcctctcccatcc 1433
 DB 474 TAAAGTTAAATGCTTAGCCTTGTACTGAGGCTGTATACAGGCACACAGCCTCTCCCATCC 415
 QY 1434 ctccagccttatctgtcatcaccatcaacccctcccatnysaccctaaacaaatcctaa 1493
 DB 414 CTCACGCTTATCTGTGCATCACCATCAACCCCTCCCATG-CACCTAAACAAATCTAACT 356
 QY 1494 tgtaattccttgaaacatgtcaggncatatacttctctctgcctgagaagctcttccct 1553
 DB 355 TGTAAATCTCTGAACATGTCAGG-CATACATTAATTCCTTCTGCTGAGAGCTCTTCCT 297
 QY 1554 gtctcttaantctagaatgatgtaaagtttgataaagttgactatcttactctatgcaa 1613
 DB 296 GTCTCTTAAATCTAGAAATGATGTAAGTTTGAATAAGTTGACTATCTTACTTCATGCAA 237
 QY 1614 agaaaggacacatatagagattcatcatcatcagacagacacaaataactaaagttaatt 1673
 DB 236 AGAAGGACACATATGAGATTTCATCATCATGAGACACAAATACATAAAGTGAATTT 177
 QY 1674 gattataagagtttagataaataatataatgaatgaacagkccacagaggaagtgttatg 1733
 DB 176 GATTATAAGAGTTTAGATAAATATATGAATGCAAGAGCCACAGAGGGAATGTTTATGGG 117
 QY 1734 gcacgtttgtaagcctggaggtgtgaagmaaaggcaggggaacctcatagtattatataa 1793
 DB 116 GCACGTTTCTAAGCCTGGGATGTGTAAGCAAGGACGAGGAACTCATAGTATCTTATATAA 57

Qy	1794	tatacttcaattctctatctatcaataatcaacaagaagttttccagaattc	1848
Db	56	TATATTCTATCTCTATCTCTATCACAATATCCAACAAGCTTTTCAGAGAATTC	2
RESULT 2			
LOCUS	AI557225	856 bp	mRNA
DEFINITION	PT2.1_15_B05.r tumor2 Homo sapiens cDNA 3', mRNA sequence.	EST	09-AUG-1999
ACCESSION	AI557225		
VERSION	AI557225.1	GI:4489588	
KEYWORDS	EST		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Huang, G.M., Ng, W.L., Farkas, J., He, L., Liang, H.A., Gordon, D., Yu, J. and Hood, L.		
TITLE	Prostate cancer expression profiling by cDNA sequencing analysis		
JOURNAL	Genomics 59 (2), 178-186 (1999)		
MEDLINE	99339982		
COMMENT	Contact: Guyang Matthew Huang University of Washington Department of Molecular Biotechnology, Box 357730, University of Washington, Seattle, WA 98195 Tel: 5106280100 Fax: 5106280108 Email: huanggm@yahoo.com.		
FEATURES	Location/Qualifiers		
source	1..856		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone_lib="tumor2"		
	/note="Organ: Prostate; Vector: pBluescript; Directional cDNA library was constructed using Lambda ZP II kit (Stratagene). mRNA was extracted from a frozen prostate tumor tissue (Mayo Clinics)."		
BASE COUNT	237 a 191 c 180 g 214 t 34 others		
ORIGIN			
Query Match	25.18;	Score 512.2;	DB 21; Length 856;
Best Local Similarity	96.3%;	Pred. No. 2.4e-126;	
Matches	570;	Conservative 4;	Mismatches 12; Indels 6; Gaps 5;
Qy	1330	actaagtcctttatccctccccttggttgattttttccagataaagttaaaatgct	1389
Db	18	ACGTAGCCTTTAATCCCTCCCTTTGTTGATTTTTTTCAGTATAAGTTAAATGCT	77
Qy	1390	tagccttgactaggctgtgtatcacg-cacagcctctcccacccctccagccttatctg	1448
Db	78	TAGCCTTGTTACTGAGGTGTATACAGGCACAGCCTCTCCCATCCTCCAGCTTATCTG	137
Qy	1449	tcataccatcaaccctcccatnysacctaaacaaaatactaaacttgtaactcttgtaac	1508
Db	138	TCATCACCATCAACCCCTCCCATG-CACCTAAACAAAATCTAAGTTGTAATTCCTTGAAC	196
Qy	1509	atgtcagncatacatrttctctctgctgagagctcttctctgtctcttaantctag	1568
Db	197	ATGTCAGG-CATACATATTCTCTGCTGCTGAGAAGCTCTTCTTGCTCTTAATCTAG	255
Qy	1569	aatgatgtaaagtttgaataagttgactactcttactctatgcgaagaagggacacatat	1628
Db	256	AATGATGTAAAGTTTGTAAATAGTTGACTACTTACTTTCATGTCGAAGAAGGCACACAT	315
Qy	1629	gagattcatcatcacatgaagacagacaaatactaaaagttaatttgattataagagttta	1688
Db	316	GAGATTCTATCATCATGAGACACAAATCTAAAGTGTAAATTGGATTAAAGAGTTTA	375
Qy	1689	gataaatatagaaatgcagaakccacagaggggaatgtttatgggggcagctttgttaagcc	1748

```
QY 1293 aatttaattacattttgttttccagtgcaaaagtgactaaagtccttttattcctccccc 1352
|||||
Db 441 AATTAAATTACATAATTTGTTTCCAGTGCAGAGTACTAAGTCTTTATCCCTCCCT 382
|||||
QY 1353 ttgtttgattttttccagtaaaagttaaatccttagcctttagctagagctgtata 1412
|||||
Db 381 TTGTTTGATTTTTTCCAGTATAAAGTTAAATGCTTAGCCTTGACTGAGGCTGTATA 322
|||||
QY 1413 cag-cacagcctcccccacccctccagccttctctgcatcaccatcaacccctcccat 1471
|||||
Db 321 CAGCCACAGCCCTCTCCCATCCCTCAGCCTTATCTGTCATCACCATCAACCCCTCCCAT 262
|||||
QY 1472 nysacctaaacaaaacttaacttgaattccttgtaactgtcaggnacatacattttctc 1531
|||||
Db 261 G-CACCTAAACAAATCTAACTTGTAAATCTTGAACATGTCAGG-CATACATATTCTCT 204
|||||
QY 1532 tctgctgagaagccttctctctctcttaantctagaatgatgaaagtgttgaataag 1591
|||||
Db 203 TCTGCTCAGAAAGCTCTCTCTGCTCTTAAATCTAGAATGATGTAAGTTTGAATAAG 144
|||||
QY 1592 ttgactacttctactcatgaagaagagacacatgatgagattcatcatcagatgagaca 1651
|||||
Db 143 TTGACTATCTTACTTCATGCAAGAAGGACACATATGAGATTCAATCATCATGAGACA 84
|||||
QY 1652 gcaataactaaaagttaattgatttaagaggttttagataaataatgaaatgcaagak 1711
|||||
Db 83 GCAATATACCCAAAGTGTAATTTGATTAAAGAGTTTAGATAAATATATGAATGCAAGAG 24
|||||
QY 1712 ccacagaggggaatgtttatggg 1733
|||||
Db 23 CCACAGAGGGGAATGTTATGGG 2
|||||

RESULT 4
AA578773 402 bp mRNA EST 12-SEP-1997
LOCUS nh24a04.s1 NCI_CGAP_Prl Homo sapiens cDNA clone IMAGE:953262, mRNA
sequence.
ACCESSION AA578773
VERSION AA578773.1 GI:2356957
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 402)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D.
, Michael Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.lnl.gov/bbrp/image/image.html
Insert Length: 565 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 373.
Location/Qualifiers
1. 402
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:953262"
/clone_lib="NCI_CGAP_Prl"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
```

/note="Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA was primed with oligo(dT)17 on 50 ng of DNase-treated, total cellular RNA obtained from 5,000-10,000 microdissected, histologically normal prostate epithelial cells. Double-stranded cDNA was ligated to EcoRI adaptors, 5 cycles of PCR applied to the cDNA with an adaptor-specific primer, and the resulting PCR product subcloned into pAMP10 by the UDG-cloning method (Life Technologies). Average insert size is 600 bp. NOTE: Not directionally cloned. This library was constructed by David Krizman."

BASE COUNT 128 a 86 c 84 g 104 t
ORIGIN

Query Match 19.3%; Score 394; DB 9; Length 402;

Best Local Similarity 98.8%; Pred. No. 1e-94;

Matches 397; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 560 ggatggggcaggaacacagatcctgttggtgatatatttgaacgggattacagatttga 619

|||||

Db 1 GGTCCAGCAGGAACACAGATCCTGTGTGATATATATTTGAACGGGATTACAGATTGA 60

|||||

QY 620 aatgaagtccacaaagtgcattaccatgagagagaacacagagaaaatcttgatgg 679

|||||

Db 61 AATGAAGTCACAAAGTGAGCATTTACCAATGAGAGGAAAAACAGCAGAGAAAATCTTGATGG 120

|||||

QY 680 ctccacagacatgac 739

|||||

Db 121 CTTCAAGACATGCAACAAACAAATGGAATGATGTGATGATGATGATGATGATGATGATG 180

|||||

QY 740 gggagagataaccagggcagagggcagaggtcagagattcgtgacctgctgacctgacctg 799

|||||

Db 181 GGGAGGAGATAACCCAGGGGAGAGGGTCAGGATTCGGCCCTGCTCCCTAAACTGTGGC 240

|||||

QY 800 ttataaccacaaatcttctcatatttctaaacctcaacacacacacacacacacacacacac 859

|||||

Db 241 TTCAATACCAAAATCATTTTCATATTTCTAAACCTCAAAACAAAGCTGTGTTGTAATATCTGAT 300

|||||

QY 860 ctctcaggttctctctggggccacacattctccatatatccacccacacacacacacacacac 919

|||||

Db 301 CTCTACGGTTCCTTCTGGGCCCAACATTCCTCATATATATCCAGCCACACTCATTTTAAATA 360

|||||

QY 920 tttagtccagatctgactgtgaccttctacacgtaga 961

|||||

Db 361 TTTAGTCCACATCTGACTGTGACCTTCTACACTGTAGA 402

|||||

RESULT 5

AA206972/3 394 bp DNA GSS 17-SEP-1998

LOCUS HS_3238_B1_G11_MR CIT Approved Human Genomic Sperm Library D Homo

sapiens genomic clone Plate=3238 Col=21 Row=N, DNA sequence.

ACCESSION AA206972

VERSION AA206972.1 GI:3617542

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 394)

REFERENCE Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,

Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and

Hood,L.

Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

JOURNAL 99380589

MEDLINE

COMMENT Contact: Mahairas GG, Wallace JC, Hood L

High throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL MEDLINE COMMENT Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR0&t2=MR0-FT0175-310800-106-b09&t3=2000-08-31&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 290.
Location/Qualifiers
1. .290
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="FT0175"
/dev_stage="Adult"
/note="Organ: prostate_tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
78 a 73 c 53 g 86 t

FEATURES source
1. .282
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="FT0175"
/dev_stage="Adult"
/note="Organ: prostate_tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
85 a 51 c 70 g 76 t

BASE COUNT 85 a 51 c 70 g 76 t
ORIGIN

Query Match 12.8%; Score 260.4; DB 147; Length 290;
Best Local Similarity 97.8%; Pred. No. 6.4e-59;
Matches 264; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 740 gggaggagataaacacgggggcagagggtcaggattctggccctgctgcctaaactgtgcg 799
Db 21 gggaggagataaacacgggggcagagggtcaggattctggccctgctgcctaaactgtgcg 80
Qy 800 ttcataacccaatcatttcattatttctaacctcaaaacaaagctgttgtaatactgat 859
Db 81 TTCATAACCAATCATTTTCATATTTCTAACCTCAAACAAAGCTGTTGTAATATCTGAT 140
Qy 860 ctctacggttcctcttgggcccaacattctccatatatccagccacacatcttttaata 919
Db 141 CTCTACGGTCTCTTCTGGGCCCAACATCTCCATATATCCAGCCACACTCATTTTAATA 200
Qy 920 ttttaqtcacagatctgactgtgacattctacacttagaataacattactatttg 979
Db 201 TTTAGTCCAGATCTGTACTGTGACCTTCTTACACTGTAGAATAACATTACTCATTTGG 260
Qy 980 ttcaagacccttcgtgtgtgccttaata 1009
Db 261 TTCAAGACCCTTCGTGGCTAACGACTA 290

RESULT 8
BF373581/c 282 bp mRNA EST 24-NOV-2000
LOCUS BF373581
DEFINITION MR0-FT0175-210800-101-d05 FT0175 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF373581
VERSION BF373581.1 GI:11335606
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (bases 1 to 282)

REFERENCE AUTHORS
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL MEDLINE COMMENT Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR0&t2=MR0-FT0175-210800-101-d05&t3=2000-08-21&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 35
High quality sequence stop: 282.
Location/Qualifiers
1. .282
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="FT0175"
/dev_stage="Adult"
/note="Organ: prostate_tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
85 a 51 c 70 g 76 t

BASE COUNT 85 a 51 c 70 g 76 t
ORIGIN

Query Match 11.1%; Score 226.8; DB 147; Length 282;
Best Local Similarity 91.3%; Pred. No. 6.6e-50;
Matches 252; Conservative 0; Mismatches 22; Indels 2; Gaps 1;

Qy 733 caagctggggaggagataaacacgggggcagagggtcaggattctggccctgctgcctaaa 792
Db 282 CAAGCTGGGGAGGAGATAAACACGGGGGAGAGGGTCAGGATTCTGGCCCTGCTGCTAAA 223
Qy 793 ctgtcgctcataacccaatcatttcattatttctaacctcaaaacaaagctgttgtaat 852
Db 222 CTGTGCGTTCATAACCAATCATTTTCATATTTCTAACCTCAAACAAAGCTGTTGTAAT 163
Qy 853 atctgatctctacggttcctcttgggcccaacattctccatatatccagccacactcatt 912
Db 162 ATCTGATCTCTACGGTTCCTCTGGGCCCAACATTTCTCCATATATCCAGCCACACTCAT 103
Qy 913 ttttaataatttagttccagatctgactgtgaccttcc--tacactgtagaataacatta 970
Db 102 TTTAATATTTAGTTCCTCCAGATCTGTACTGTGACCTTCTTCTCATCTGTGTAATAACTTA 43
Qy 971 ctcatctgttccaagacccttcgtgtgtgcctta 1006
Db 42 CTCATTTGTTCAAGACCTTCACTTGCTAGCGACTA 7

RESULT 9

```

BF858371/c      226 bp      mRNA      EST      16-JAN-2001
LOCUS          RC5-FT0193-211100-012-E11 FT0193 Homo sapiens cDNA, mRNA sequence.
DEFINITION    BF858371
ACCESSION     BF858371
VERSION       BF858371.1 GI:12246115
KEYWORDS      EST.
SOURCE        human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 226)
AUTHORS      Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
               Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
               Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
               Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
               M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
               Simpson,A.J.
TITLE        Shotgun sequencing of the human transcriptome with ORF expressed
               sequence tags
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE      20202663
COMMENT      Contact: Simpson A.J.G.
               Laboratory of Cancer Genetics
               Ludwig Institute for Cancer Research
               Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
               Brazil
               Tel: +55-11-2704922
               Fax: +55-11-2707001
               Email: asimpson@ludwig.org.br
               This sequence was derived from the FAPESP/LICR Human Cancer Genome
               Project. This entry can be seen in the following URL
               (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=RC5&t2=RC5-FT0193-
               211100-012-E11&t3=2000-11-21&t4=1)
               Seq primer: puc 18 forward
               High quality sequence stop: 226.
               Location/Qualifiers
                 1..226
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /clone_lib="FT0193"
                   /dev_stage="Adult"
                   /note="Organ: prostate_tumor; Vector: puc18; Site_1: SmaI;
                   Site_2: SmaI; A mini-library was made by cloning products
                   derived from ORESTES PCR (U.S. Letters Patent application
                   No. 196,716 - Ludwig Institute for Cancer Research)
                   profiles into the pUC 18 vector. Reverse transcription of
                   tissue mRNA and cDNA amplification were performed under
                   low stringency conditions."
BASE COUNT    79 a      29 c      59 g      59 t
ORIGIN
Query Match   10.5%; Score 214.4; DB 170; Length 226;
Best Local Similarity 99.5%; Pred. No. 1.3e-46;
Matches 215; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  787 cctaaactgtgggttcataccaaatcatttcatttccatatttccacccctcaaaaacaaagctgt 846
Db  226 CCTAAACTGGGTTCATACCAAAATCATTTTCATATTTCTAACCCCTCAAAACAAAGCTGT 167

QY  847 tgaatatctgatctctacgggttcctctctggggcccaacattctccatatatccagccaca 906
Db  166 TGAATATCTGATCTCTACGGTTCCTTCCTGGGCCCAACATTTCTCCATATATCCAGCCACA 107

QY  907 ctcatttttaattagttccagatctgtactgtgacctttctacactgtagaataac 966
Db  106 CCAATTTTAAATTTAGTTCCAGATCTGTACTGTGACCTTTCTACACTGTAGATAAC 47

QY  967 attactcattttgttcaaaagaccttcgtgttgctg 1002
Db  46 ATTACTCATTTTGTTCAAAGACCCCTCGTGTGCAG 11

```

```

RESULT 10
BF373406      167 bp      mRNA      EST      24-NOV-2000
LOCUS          IL2-FT0159-070800-120-H01 FT0159 Homo sapiens cDNA, mRNA sequence.
DEFINITION    BF373406
ACCESSION     BF373406
VERSION       BF373406.1 GI:11335431
KEYWORDS      EST.
SOURCE        human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 167)
AUTHORS      Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
               Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
               Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
               Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
               M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
               Simpson,A.J.
TITLE        Shotgun sequencing of the human transcriptome with ORF expressed
               sequence tags
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE      20202663
COMMENT      Contact: Simpson A.J.G.
               Laboratory of Cancer Genetics
               Ludwig Institute for Cancer Research
               Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
               Brazil
               Tel: +55-11-2704922
               Fax: +55-11-2707001
               Email: asimpson@ludwig.org.br
               This sequence was derived from the FAPESP/LICR Human Cancer Genome
               Project. This entry can be seen in the following URL
               (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=IL2&t2=IL2-FT0159-
               070800-120-H01&t3=2000-08-07&t4=1)
               Seq primer: puc 18 forward
               High quality sequence stop: 167.
               Location/Qualifiers
                 1..167
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /clone_lib="FT0159"
                   /dev_stage="Adult"
                   /note="Organ: prostate_tumor; Vector: puc18; Site_1: SmaI;
                   Site_2: SmaI; A mini-library was made by cloning products
                   derived from ORESTES PCR (U.S. Letters Patent application
                   No. 196,716 - Ludwig Institute for Cancer Research)
                   profiles into the pUC 18 vector. Reverse transcription of
                   tissue mRNA and cDNA amplification were performed under
                   low stringency conditions."
BASE COUNT    47 a      38 c      49 g      33 t
ORIGIN
Query Match   8.2%; Score 167; DB 147; Length 167;
Best Local Similarity 100.0%; Pred. No. 6.3e-34;
Matches 167; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  419 cacagaagcacaagaagacacagatccctggagaaatgccggcccccattcttg 478
Db  1 CACAGAAGCACAAGAAGGAAGCACAGATCCTCTGGGAGAAATGCCGGCCGCCATCTTG 60

QY  479 ggatcatcgatgagcctcgccctgtgcctgtgctcccgcttgtgaggaagacattagaaaa 538
Db  61 GGTATCATGATGAGCCTCGCCCTGTGCTGTGCTCCCTGTGAGGGAAGACATTAGAAAA 120

QY  539 tgaattgatgttctcttaaaagatggcgaggaaacagatcctgtt 585
Db  121 TGAATTGATGTCTCTTAAAGGATGGGCAGGAAACACAGATCCTGTT 167

RESULT 11
AQ246715      451 bp      DNA      GSS      06-OCT-1998
LOCUS          AQ246715

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 02:27:18 ; Search time 176.8 Seconds
(without alignments)
2181.147 Million cell updates/sec

Title: US-09-402-713A-1

Perfect score: 2037

Sequence: 1 agaagctggcatcagaaaaa.....caataaagaatttacaaga 2037

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA.*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PT05_COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	59.8	2.9	7218	1	US-08-232-463-14
2	57.4	2.8	7218	1	US-08-232-463-14
3	47	2.3	10968	2	US-08-680-327-2
4	47	2.3	10968	4	US-09-228-246-1
5	45.4	2.2	5134	2	US-08-310-912A-157
6	45.4	2.2	5134	5	PCT-US95-04589-157
7	45.4	2.2	5475	2	US-08-680-327-1
8	45.4	2.2	5475	4	US-09-228-246-3
9	35.8	1.8	14507	3	US-08-785-150-1
10	35	1.7	4129	2	US-08-370-319C-12
11	35	1.7	4129	4	US-09-224-834-12
12	34.8	1.7	5718	3	US-08-714-918-48
13	34.8	1.7	5718	4	US-09-265-315-48
14	34.8	1.7	5718	4	US-09-265-315-48
15	34.8	1.7	5718	4	US-09-265-315-48
16	34.2	1.7	2679	1	US-07-977-434-11
17	34.2	1.7	2679	1	US-08-458-819-11
18	34.2	1.7	2679	5	PCT-US91-07035-11
19	33.8	1.7	2570	2	US-09-056-075-2
20	33.8	1.7	246240	2	US-08-724-394A-20
21	33.8	1.7	246240	2	US-08-724-394A-21
22	33.8	1.7	246240	2	US-08-724-394A-22
23	33.6	1.6	3084	1	US-08-551-437-3
24	33.6	1.6	3084	3	US-09-004-225-3
25	33.6	1.6	3084	3	US-09-084-346-3
26	33.2	1.6	5173	1	US-08-242-677-1
27	33	1.6	1952	1	US-08-333-358-1

28	33	1.6	1952	1	US-08-463-694-1	Sequence 1, Appli
29	33	1.6	1952	1	US-08-694-501-1	Sequence 1, Appli
30	33	1.6	2950	5	PCT-US93-08386-7	Sequence 7, Appli
31	33	1.6	2968	5	PCT-US93-08386-1	Sequence 1, Appli
32	33	1.6	8920	2	US-08-446-855A-1	Sequence 1, Appli
33	33	1.6	8920	4	US-09-150-741-1	Sequence 1, Appli
34	33	1.6	720	4	US-08-446-935-4	Sequence 4, Appli
35	32.8	1.6	2395	4	US-08-446-935-7	Sequence 7, Appli
36	32.8	1.6	2405	1	US-08-454-097-30	Sequence 30, Appli
37	32.8	1.6	2405	3	US-08-185-359-30	Sequence 30, Appli
38	32.8	1.6	5356	4	US-08-446-935-1	Sequence 1, Appli
39	32.6	1.6	1400	1	US-07-989-845-1	Sequence 1, Appli
40	32.6	1.6	1400	5	PCT-US93-11298-1	Sequence 1, Appli
41	32.6	1.6	3607	1	US-08-647-351B-1	Sequence 1, Appli
42	32.6	1.6	246240	2	US-08-724-394A-20	Sequence 20, Appli
43	32.6	1.6	246240	2	US-08-724-394A-21	Sequence 21, Appli
44	32.6	1.6	246240	2	US-08-724-394A-22	Sequence 22, Appli
45	32	1.6	713	3	US-08-532-896-26	Sequence 26, Appli

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PT99pt-Fls
; US-08-232-463-14

```
Query Match 2.9%; Score 59.8; DB 1; Length 7218;
Best Local Similarity 5.1%; Pred. No. 1.2e-07;
Matches 20; Conservative 219; Mismatches 155; Indels 0; Gaps 0;

QY 1178 ttcattcccaactttgtgccattctcaagacacctcaaatgtcattccattcaataatcac 1237
Db 1076 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1135
QY 1238 aggattaaactttttttaacctgaagaattcaatgtttacatgcagctatgggaattt 1297
Db 1136 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1195
QY 1298 aattacataatttggtttcocagtcgaagatgactaagtcctttatccctcccttggtt 1357
Db 1196 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1255
QY 1358 tgattttttccagtataaagttaaaagtcttagctgtactgtactgagctgtatcacga 1417
Db 1256 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1315
QY 1418 cagcctctcccatccctccagcttctctgtctcatcaccatcaacccctcccatnysacc 1477
Db 1316 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1375
QY 1478 taacaaaatctaacttgaattccttgaaatgacatgtcaggnacatactttctctgcc 1537
Db 1376 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1435
QY 1538 tgagaagctctctctgtctctcttaantctagaat 1571
Db 1436 TACCAAAATCTTCTATCTCTTTAACTACTGTCAT 1469
```

```
RESULT 2
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22131-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgt-Fls
US-08-232-463-14

Query Match 2.8%; Score 57.4; DB 1; Length 7218;
Best Local Similarity 4.3%; Pred. No. 6.4e-07;
Matches 16; Conservative 211; Mismatches 142; Indels 0; Gaps 0;

QY 422 aggaagacaaagaagcacagagatccctgggagaaatgccggccgccatcttgggt 481
Db 1407 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1348
QY 482 catcgatgagctgcctgcctgtcctgtgctgctgtgaggaagacattagaaaatga 541
Db 1347 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1288
QY 542 attgatgttctctaaagatggcaggagaaacagatcctgtgtgatatatttga 601
Db 1287 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1228
QY 602 acgggattcacagattgaaatgaagtcaacaagtgcattaccatgagagaaacag 661
Db 1227 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1168
QY 662 acgagaaaaatcttgatggcttcacaagacatgcacaaacaaatggaatactgtgata 721
Db 1167 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1108
QY 722 catgagggcgcaagctggggaggagataaacacaggggcagaggggcagattctggccc 781
Db 1107 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1048
QY 782 tgctgccta 790
Db 1047 TGCAGCCAA 1039
```

```
RESULT 3
US-08-680-327-2/c
; Sequence 2, Application US/08680327
; Patent No. 5859321
; GENERAL INFORMATION:
; APPLICANT: Staskawicz, Brian S., Oldroyd, Giles Edward,
; APPLICANT: Salmeron, John M., Rommens, Caius
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PLANT
; TITLE OF INVENTION: PATHOGEN RESISTANCE
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klarquist Sparkman Campbell Leigh &
; ADDRESSEE: Whinston
; STREET: One World Trade Center
; STREET: 121 S.W. Salmon Street
; STREET: Suite 1600
; CITY: Portland
; STATE: Oregon
; COUNTRY: United States of America
; ZIP: 97204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disk, 3-1/2 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/680,327
; FILING DATE: July 11, 1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
```



```

1  FILING DATE:
2  CLASSIFICATION:
3  PRIOR APPLICATION DATA:
4  APPLICATION NUMBER: US/08/586,509
5  FILING DATE: 11-JAN-96
6  ATTORNEY/AGENT INFORMATION:
7  NAME: Perkins, Patricia Anne
8  REGISTRATION NUMBER: 34,693
9  REFERENCE/DOCKET NUMBER: 2841
10 TELECOMMUNICATION INFORMATION:
11 TELEPHONE: (206)587-0430
12 TELEFAX: (206)233-0644
13 INFORMATION FOR SEQ ID NO: 1:
14 SEQUENCE CHARACTERISTICS:
15 LENGTH: 14507 base pairs
16 TYPE: nucleic acid
17 STRANDEDNESS: single
18 TOPOLOGY: NO. 6027915 Relevant
19 MOLECULE TYPE: DNA (genomic)
20 HYPOTHETICAL: NO
21 ANTI-SENSE: NO
22 ORIGINAL SOURCE:
23 ORGANISM: Chinese hamster
24 IMMEDIATE SOURCE:
25 CLONE: 2A5-3 lambda CHO sequence
26
27 US-08-785-150-1

```

	Query Match	1.8%	Score 35.8;	DB 3;	Length 14507;
	Best Local Similarity	46.9%;	Pred. No. 3.1;		
Matches 112;	Conservative 0;	Mismatches 127;	Indels 0;	Gaps 0;	
QY	1035	gtttctgccccggagatctgtgaacaggctcgggaagcatctcaagaatcttccagggt	1094		
Db	10959	GGTTCCTCCACAGTGGAATGATGTCAACAAGGAAATTATTGCAGCCTGGTGCTAGTG	10900		
QY	1095	tatacttagcacacagcatgatcattacgagtgtaatttatcaaacatcatcct	1154		
Db	10899	GCTACAATCTPAGGCCGGCTTTAGAGGCAGCACGCTTCTTAANTCAGGTCCTGTGTGTAC	10840		
QY	1155	cagtgccttcgcccataactaaaattcctcccacttttgtcccactctccaagacctoa	1214		
Db	10839	GACGATGACTAAGCGCCAAGTATAACAGGGCAACTGCTTTGAGAATAATTAAAGTGTCA	10780		
QY	1215	aaatgctatccattaatatcacaggataaacttttttttaacctcggaagaattcaa	1273		
Db	10779	CAACGTAGTTTTAGTAGTACAGACAGCAAACATTTATTTTATTTTCGAAAAATATAA	10721		

RESULT 10
 US-08-319c-12/c
 ; Sequence 12, Application US/08370319C
 ; Patent No. 5856091
 ; GENERAL INFORMATION:
 ; APPLICANT: Brichard, Vincent; Van Pel, Aline;
 ; APPLICANT: Traversari, Catia; W Irel, Thomas; Coullie, Pierre;
 ; APPLICANT: Boon-Fallieur, Thierry; De plaen, Etienne
 ; TITLE OF INVENTION: ISOLATED NUCLEIC ACID SEQUENCE CODING FOR A
 ; TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PRECURSOR PROCESSED TO AT LEAST ONE TU
 ; TITLE OF INVENTION: REJECTION ANTIGEN PRESENTED BY HLA-A2
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Felfe & Lynch
 ; STREET: 805 Third Avenue
 ; CITY: New York City
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10022
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
 ; COMPUTER: IBM PS/2
 ; OPERATING SYSTEM: PC-DOS
 ; SOFTWARE: Wordperfect

; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/370,319C
 ; FILING DATE: 10-JANUARY-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/272,351
 ; FILING DATE: 8-JULY-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/032,978
 ; FILING DATE: 18-MAR-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Hanson, No. 5856091man D.
 ; REGISTRATION NUMBER: 30,946
 ; REFERENCE/DOCKET NUMBER: LUD 5377.1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 688-9200
 ; TELEFAX: (212) 838-3884
 ; INFORMATION FOR SEQ ID NO: 12:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4129 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; FEATURE:
 ; OTHER INFORMATION: The sequence is preceded by an
 ; OTHER INFORMATION: unsequenced portion of from 4.7 to 5.3
 ; OTHER INFORMATION: kilobases
 ; US-08-370-319C-12

Query Match	1.7%	Score 35	DB 2	Length 4129	
Best Local Similarity	48.7%	Pred. No. 2.5			
Matches	95	Conservative 0	Mismatches 100	Indels 0	Gaps 0
Qy 1190	ttttgtcccatctc	caagaccctcaaaatgctcattccattaatatcacaggattaaacttt	1249		
Db 2703	TCITTTCTACATAC	TATATTTTTTGAATATGATATGTATTTATATACATGCTTATTTC	2644		
Qy 1250	tttttttaacctggaagaattc	aattgattacatgcagctatgggaatttaattacatattt	1309		
Db 2643	AACTCCCTGTGTTT	TAAGTGTTCATAGCACCATGTGACTATGGCTAAGCCATCAGACAG	2584		
Qy 1310	tgtttccagtggaagatgactaa	gcttcttatccctccctctgttttgatttttttc	1369		
Db 2583	CACAGCTCTATACAA	TAAATAAATGAGTACITTTATCCCTTCCTTGTATATATCTGCATC	2524		
Qy 1370	cagtataaagttaa	1384			
Db 2523	CATTAAAAAGGCAA	2509			

RESULT 11
US-09-224-834-12/c
; Sequence 12, Application US/09224834
; Patent No. 620111
; GENERAL INFORMATION:
; APPLICANT: Brichard, Vincent; Van Pel, Aline;
; APPLICANT: Traversari, Catia; Wifel, Thomas; Pierre;
; APPLICANT: Boon-Fallieur, Thierry; De Plaen, Etienne
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID SEQUENCE CODING FOR A
; TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PERCUSOR CODED FOR A
; TITLE OF INVENTION: REJECTION ANTIGEN PRESENTED BY HLA-A2
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM ps/2

```

: OPERATING SYSTEM: PC-DOS
: SOFTWARE: wordperfect
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/224,834
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/370,319
: FILING DATE:
: PRIOR APPLICATION DATA:
: FILING DATE: 18-MAR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Hanson, No. 6201111man D.
: REGISTRATION NUMBER: 30,946
: REFERENCE/DOCKET NUMBER: LUD 5377
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 688-9200
: TELEFAX: (212) 838-3884
: INFORMATION FOR SEQ ID NO: 12:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4129 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: FEATURE:
: OTHER INFORMATION: The sequence information
: OTHER INFORMATION: unsequenced portions
: OTHER INFORMATION: kilobases
: US-09-224-834-12

```

Query Match	1.7%	Score 35	DB 4	Length 4129
Best Local Similarity	48.7%	Pred. No. 2.5		
Matches 95	Conservative 0	Mismatches 100	Indels 0	Gaps 0
Qy 1190	ttttgtgccattctcaagacctcaaaatgcattccatttaataatcacagattaaacttt	1249		
Db 2703	TCTTTTTCACATACATATATTTTTTGAATATGATATGTATTATATACATGCTTATTTC	2644		
Qy 1250	tttttttaacctgggaagaattcaatgtttacatgcagctatgggaatttaattacatattt	1309		
Db 2643	AACCTCTGTGTTTTAAAGTGTTCATAGCACCATGTGACTAGTGGCTAAGCCATCAGACAG	2584		
Qy 1310	tgttttcagtgcaaaagactgaactaagtcctttatcccttccctttgttgtatttttttc	1369		
Db 2583	CACAGCTCTATACATAAATAATAGGTACTTTATCCCTCCCTCTGTATATATCTGCATC	2524		
Qy 1370	cagtataaagtgtaaa	1384		
Db 2523	CATTAAAAAGGCAAA	2509		

```

RESULT 12
US-08-714-918-48
: Sequence 48, Application US/08714918
: Patent No. 6037123
: GENERAL INFORMATION:
: APPLICANT: Benton, Bret
: APPLICANT: Lee, Ving
: APPLICANT: Malouin, Francois
: APPLICANT: Martind, Patrick K.
: APPLICANT: Schmid, Molly B.
: APPLICANT: Sun, Dongxu
: TITLE OF INVENTION: STAPHYLOCOCC
: TITLE OF INVENTION: TARGET GENES
: NUMBER OF SEQUENCES: 111
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Lyon & Lyon
: STREET: 633 West Fifth Street
: STREET: Suite 4700
: CITY: Los Angeles
: STATE: California

```

COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
MEDIUM TYPE: Storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/714,918
FILING DATE: September 13, 1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/009,102
FILING DATE: December 22, 1995
APPLICATION NUMBER: 60/003,798
FILING DATE: September 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 222/005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ. ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 5718 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-714-918-48

	Query Match	1.7%	Score 34.8	DB 3	Length 5718
	Best Local Similarity	48.1%	Pred. No. 3.5		
	Matches 99	Conservative 0	Mismatches 107	Indels 0	Gaps 0
Qy	1198	ccattctcaagacactcaaaatgtcattccattcaatatcacaggatgaactttttttta	1257		
Db	283	CCATCAGTAATAATGCACAGTATTTCTTTAATGTGTAATCATATACATTTCTTTTA	342		
Qy	1258	acctggaagaattccaatggtttacatgcagctatgggaatttaattacacattttgtttcc	1317		
Db	343	TCTTCATAAAATATCTTTACACGGTTTCAGCAACGCCAGGTGAGTAGCTAAACTTAAATCC	402		
Qy	1318	agtgcaaaagatgcactaagtctcttattccctccctctgtttgattttttttccagataaa	1377		
Db	403	TCTTTATTAGTAACCTTTTACATTTGGTTTAACTTCTTAATTTACCTTGATTAAGTTTGTC	462		
Qy	1378	agttaaaatgccttagcctttgactga	1403		
Db	463	ATTTCCTCAATGCTTCATCTCTTAAATGA	488		

RESULT 13
US-09-265-315-48
; Sequence 48, Application US/09265315
; Patent No. 6187541
; GENERAL INFORMATION:
; APPLICANT: Benton, Bret
; APPLICANT: Lee, Ving J.
; APPLICANT: Malouin, Francois
; APPLICANT: Martin, Patrick K.
; APPLICANT: Schmid, Molly B.
; APPLICANT: Sun, Dongxu
; TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
; TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
; TITLE OF INVENTION: TARGET GENES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street

Search completed: July 31, 2001, 07:18:22
Job time: 17464 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 02:29:48 ; Search time 478.87 Seconds
(without alignments)
2670.945 Million cell updates/sec

Title: US-09-402-713a-1
Perfect score: 2037
Sequence: 1 agaagctggcatcagaaaa.....caataaagaatttacaaga 2037

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601.*
1: /SID88/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2032	99.8	2037	19	Prostate cancer an
2	2030.4	99.7	3582	19	Prostate cancer an
3	1772.4	87.0	1872	19	Prostate cancer an
4	1737.8	85.3	3112	21	Human immunogenic
5	1731.4	85.0	2426	21	Human immunogenic
6	1712.4	84.1	2229	21	Human immunogenic
7	725.4	35.6	820	19	Prostate cancer an
8	722.4	35.5	812	21	Human immunogenic
9	524	25.7	597	20	Human secreted pro
10	457.2	22.4	718	21	Human immunogenic
11	319.6	15.7	437	21	Human secreted pro

c	12	288.4	14.2	301	21	AAAO6520	Human immunogenic
c	13	235.4	11.6	359	20	AAZ33445	Human prostate can
c	14	136	6.7	936	22	AAF58252	Oligonucleotide D1
c	15	136	6.7	936	22	AAF58254	Oligonucleotide D1
c	16	136	6.7	936	22	AAF58257	Oligonucleotide D1
c	17	136	6.7	936	22	AAF58259	Oligonucleotide D2
c	18	136	6.7	936	22	AAF58262	Oligonucleotide D2
c	19	136	6.7	936	22	AAF58265	Oligonucleotide D1
c	20	134.4	6.6	936	22	AAF58252	Oligonucleotide D1
c	21	134.4	6.6	936	22	AAF58254	Oligonucleotide D1
c	22	134.4	6.6	936	22	AAF58257	Oligonucleotide D1
c	23	134.4	6.6	936	22	AAF58259	Oligonucleotide D2
c	24	134.4	6.6	936	22	AAF58262	Oligonucleotide D2
c	25	134.4	6.6	938	22	AAF58255	Oligonucleotide D1
c	26	99	4.9	123	19	AAV33791	Prostate cancer an
c	27	50.8	2.5	244	22	AAF58238	Oligonucleotide D1
c	28	47.6	2.3	244	22	AAF58238	Oligonucleotide D1
c	29	47	2.3	10968	19	AAV17789	Tomato Prf genomic
c	30	45.4	2.2	5475	19	AAV17777	Tomato Prf cDNA.
c	31	40	2.0	796	20	AAV37484	Human secreted pro
c	32	39	1.9	2206	21	AAV12390	Human interferon-b
c	33	39	1.9	10325	20	AAV76516	Plasmid containing
c	34	37.6	1.8	1297	13	AAQ21645	3' coding sequence
c	35	37.6	1.8	1770	14	AAQ37684	P.falciparum antig
c	36	37.6	1.8	1770	14	AAQ37124	P.falciparum antig
c	37	37.4	1.8	1255	22	AAF72762	Human prostate can
c	38	37.4	1.8	1325	21	AAF15730	Human prostate can
c	39	37.4	1.8	1325	22	AAF72787	Human prostate can
c	40	37.2	1.8	1751	21	AAC68122	Human secreted pro
c	41	36.8	1.8	5804	18	AAV74560	Staphylococcus aur
c	42	36.4	1.8	2643	15	AAQ66096	S. cerevisiae Plc1
c	43	36.2	1.8	1451	21	AAF21739	Human breast and o
c	44	36.2	1.8	1517	21	AAC34792	Arabidopsis thalia
c	45	36.2	1.8	1630	20	AAV13310	Enterococcus faeca

ALIGNMENTS

RESULT 1
AAV62427
ID AAV62427 standard; cDNA; 2037 BP.
XX
AC AAV62427;
XX
DT 30-DEC-1998 (first entry)
XX
DE Prostate cancer antigen (PCA3) cDNA splice variant 1.
XX
KW Prostate cancer antigen cDNA splice variant 1.
KW PC; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 379..534
FT /tag= a
FT /product= "PCA3 protein variant 1"
FT polyA_signal 2019..2024
FT /tag= b
XX
PN WO9845420-A1.
XX
PD 15-OCT-1998.
XX
PF 09-APR-1998; 98WO-CA00346.
XX
PR 10-APR-1997; 97US-0041836.
XX
PA (DIAG-) DIAGNOCURE INC.
XX
PI Bussemakers MJG;

DR WPI: 1998-568347/48.
XX P-PSDB; AAW79736.
PT New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
PT prevention and treatment of prostatic cancer
XX
PS Claim 3; Fig 2B-2J; 111pp; English.
XX
CC The present sequence represents the prostate cancer antigen (PCA3)
CC cDNA splice variant 1 sequence comprising of exons 1, 2, 3, 4a and
CC 4b of the PCA3 gene. The PCA3 cDNA splice variant 1 sequence,
CC isolated from a human primary prostatic tumour tissue cDNA library,
CC was found in approximately 5% of the cDNA clones isolated. The
CC invention claims for PCA3 cDNA variants and the proteins they encode.
CC The invention also claims for antibodies against PCA3 protein. The
CC antibodies are claimed to be useful for detecting PCA3 protein in
CC immunoassay tests, for diagnosing, assessing and prognosing of
CC prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin
CC or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed
CC to be useful for treating PC, while determining elevated levels of
CC PCA3 (as RNA or protein) is useful for detecting a predisposition
CC to development of PC, e.g. in prenatal tests. Detecting PCA3 protein
CC allows differentiation between malignant and benign prostatic disease,
CC and the level of PCA3 expression allows correlation with the grade of
CC tumour. PCA3 protein and its fragments are also claimed to be useful
CC in vaccines for preventing PC; in drug screens for identifying
CC specific (ant)agonists (potentially useful therapeutically) and for
CC studying protein-DNA interactions.
XX
SQ Sequence 2037 BP; 622 A; 426 C; 406 G; 575 T; 8 other;

Query Match 99.8%; Score 2032; DB 19; Length 2037;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2037; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaagctgcatcagaaaaacacagagggagattgtgtgctgcagcccgagggagaccag 60
DB 1 ageagctgcatcagaaaaacacagagggagattgtgtgctgcagcccgagggagaccag 60
QY 61 gaagatctcatgtgtgggaagacctgatgatacagaggaattacaacacataacttag 120
DB 61 gaagatctcatgtgtgggaagacctgatgatacagaggaattacaacacataacttag 120
QY 121 tgtttcaatgaacacacagataataaagtgaagagctagtcgcgtgagctccctcagt 180
DB 121 tgtttcaatgaacacacagataataaagtgaagagctagtcgcgtgagctccctcagt 180
QY 181 gacacagggctggatcaccatcgacggcactttctgagtactcagtcgacgaaagaaga 240
DB 181 gacacagggctggatcaccatcgacggcactttctgagtactcagtcgacgaaagaaga 240
QY 241 ctacagacatcctcaatggcaggggtgagaaataagaaggctgctgactttaccatctga 300
DB 241 ctacagacatcctcaatggcaggggtgagaaataagaaggctgctgactttaccatctga 300
QY 301 ggcacacatctgctgaatggagataataacatacactagaaacagcaagatgacaata 360
DB 301 ggcacacatctgctgaatggagataataacatacactagaaacagcaagatgacaata 360
QY 361 taatgtctaagttagtagactgtttttgacattttccagccctttaaatattccacacaca 420
DB 361 taatgtctaagttagtagactgtttttgacattttccagccctttaaatattccacacaca 420
QY 421 caggaagcacaagaagaaacacagagatccctgggagaaatcccgccgcacatcttggg 480
DB 421 caggaagcacaagaagaaacacagagatccctgggagaaatcccgccgcacatcttggg 480
QY 481 tcatgatgagctgcctcctgtgctcctggtcccgctgtgagggagagcattagaaaaatg 540
DB 481 tcatgatgagctgcctcctgtgctcctggtcccgctgtgagggagagcattagaaaaatg 540
QY 541 aattgatgtgttcttaaaaggatgggcagggaaaaacagatcctgttggatatttttg 600

DB 541 aattgatgtgttcttaaaaggatgggcagggaaaaacagatcctgttggatatttttg 600
QY 601 aacgggattacagatttgaatgaagtcaaaagtgaagcattaccatgagaggaaaaaca 660
DB 601 aacgggattacagatttgaatgaagtcaaaagtgaagcattaccatgagaggaaaaaca 660
QY 661 gacgagaaatcttgatggcttcacaagacatgcacaacaaacaaaatggaaactgtgatg 720
DB 661 gacgagaaatcttgatggcttcacaagacatgcacaacaaacaaaatggaaactgtgatg 720
QY 721 acatgaggcagccaagctggggagagataaacacagggggcagagggtcaggattctggcc 780
DB 721 acatgaggcagccaagctggggagagataaacacagggggcagagggtcaggattctggcc 780
QY 781 ctgctgcctcaactgtgcgtttcataacaaaatcatttcatatttcaacccccaacaaa 840
DB 781 ctgctgcctcaactgtgcgtttcataacaaaatcatttcatatttcaacccccaacaaa 840
QY 841 agctgtgtaatatctgatctctacoggttccctctggtggcccaacattctccatatcca 900
DB 841 agctgtgtaatatctgatctctacoggttccctctggtggcccaacattctccatatcca 900
QY 901 gccacactcatctttaaatttagttccacagatctgtactgtgaccttctacactgtag 960
DB 901 gccacactcatctttaaatttagttccacagatctgtactgtgaccttctacactgtag 960
QY 961 aataacattactcattttgttcaaaagacctctgttctgctgcctaataatgactgact 1020
DB 961 aataacattactcattttgttcaaaagacctctgttctgctgcctaataatgactgact 1020
QY 1021 gttttctcaaggagtgcttctgcccaggggatctgtgaacaggtcgggaagcatctcaa 1080
DB 1021 gttttctcaaggagtgcttctgcccaggggatctgtgaacaggtcgggaagcatctcaa 1080
QY 1081 gatcttccaggggtatactactactagcacacagcatgatactacggagtgaaattatcta 1140
DB 1081 gatcttccaggggtatactactactagcacacagcatgatactacggagtgaaattatcta 1140
QY 1141 atcaacatcatcctcagtgcttcttgcccaactgaaatcatttcccaacttttgcacca 1200
DB 1141 atcaacatcatcctcagtgcttcttgcccaactgaaatcatttcccaacttttgcacca 1200
QY 1201 ttctcaagacctcacaatgctatcttcatttaataatcacagattaaactttttttaacc 1260
DB 1201 ttctcaagacctcacaatgctatcttcatttaataatcacagattaaactttttttaacc 1260
QY 1261 tggagaattcaatgttacatgcagctatgggaatttaattacatatatttgttttccagt 1320
DB 1261 tggagaattcaatgttacatgcagctatgggaatttaattacatatatttgttttccagt 1320
QY 1321 gcaagatgactaaagtcccttaccctccctctgttgaattttttccagataaaagt 1380
DB 1321 gcaagatgactaaagtcccttaccctccctctgttgaattttttccagataaaagt 1380
QY 1381 taaaatgcttagcctttagtgcagctgtatcacagcacagcctctcccatccctccagc 1440
DB 1381 taaaatgcttagcctttagtgcagctgtatcacagcacagcctctcccatccctccagc 1440
QY 1441 ctatatctgataccaatcaacccctcccatnysacctaataacatacttaattgttaatt 1500
DB 1441 ctatatctgataccaatcaacccctcccatnysacctaataacatacttaattgttaatt 1500
QY 1501 ccttgaacatgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1560
DB 1501 ccttgaacatgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1560
QY 1561 aantctagaatgatgataaagtgttgaataagtgtgactatcttacttcatgcgaaagaagg 1620
DB 1561 aantctagaatgatgataaagtgttgaataagtgtgactatcttacttcatgcgaaagaagg 1620
QY 1621 acacatatgagattcatcatcatcatcatcatcatcatcatcatcatcatcatcatcatcat 1680
DB 1621 acacatatgagattcatcatcatcatcatcatcatcatcatcatcatcatcatcatcatcat 1680

Db	1621	acacatagagattcattcatcacatgagacagcaaaatactactaaaagttgtaatttgattata	1680
Qy	1681	agagtttagataaaatataatgaaatgcaagakccacagagggaaatggtttatggggcacgtt	1740
Db	1681	agagtttagataaaatataatgaaatgcaagakccacagagggaaatggtttatggggcacgtt	1740
Qy	1741	tgtaagcctggagatgtaagmmaaaggcgaggaaacctcatagatctctataataatactt	1800
Db	1741	tgtaagcctggagatgtaagmmaaaggcgaggaaacctcatagatctctataataatactt	1800
Qy	1801	catttctctatctctatcacaaatacccaacaagctttttcacagaattcatgctgagtgcaaa	1860
Db	1801	catttctctatctctatcacaaatacccaacaagctttttcacagaattcatgctgagtgcaaa	1860
Qy	1861	tcccaaaaggttaacctttatccatttcatggtgagtcgcgtttagaattttggcaaatca	1920
Db	1861	tcccaaaaggttaacctttatccatttcatggtgagtcgcgtttagaattttggcaaatca	1920
Qy	1921	tactgtgctacttatctccaactttgagatggtttgtcccttgagttaattgaaagaata	1980
Db	1921	tactgtgctacttatctccaactttgagatggtttgtcccttgagttaattgaaagaata	1980
Qy	1981	gggcactcttggagccactttagggtttcaactcctcctggcaataaagaattttacaaga	2037
Db	1981	gggcactcttggagccactttagggtttcaactcctcctggaataaagaattttacaaga	2037

RESULT	2
AAV62430	
ID	AAV62430 standard; cDNA; 3582 BP.
XX	
AC	AAV62430;
XX	
DT	30-DEC-1998 (first entry)
XX	
DE	Prostate cancer antigen (PCA3) wild-type cDNA.
XX	
KW	Prostate cancer antigen cDNA; PCA3; prostatic cancer;
KW	PC; ds.
XX	
OS	Homo sapiens.
XX	
Key	Location/Qualifiers
FH	401..556
FT	/tag= a
FT	/product= "PCA3 protein"
FT	983..987
polyA_signal	
FT	/tag= b
polyA_signal	2041..2046
FT	/tag= c
polyA_signal	2597..2602
FT	/tag= d
polyA_signal	3494..3496
FT	/tag= e
XX	
WO	9845420-A1.
PN	
XX	
PD	15-OCT-1998.
XX	
Pf	09-APR-1998; 98WO-CA00346.
XX	
PR	10-APR-1997; 97US-0041836.
XX	
PA	(DIAG-) DIAGNOCURE INC.
XX	
PI	Bussemakers MJG;
XX	
DR	WPI; 1998-568347/48.
DR	P-PSDB; AAW79738.
XX	
PT	New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
PT	prevention and treatment of prostatic cancer
XX	

Claim 3; Fig 5B-5F; 11pp; English.

The present sequence represents the prostate cancer antigen (PCA3) wild-type cDNA sequence comprising of exons 1, 2, 3, 4a-4d of the PCA3 gene. The invention claims for PCA3 cDNA variants and the proteins they encode. The invention also claims for antibodies against PCA3 protein. The antibodies are claimed to be useful for detecting PCA3 protein in immunoassay tests, for diagnosing, assessing and prognosing of prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed to be useful for treating PC, while determining elevated levels of PCA3 (as RNA or protein) is useful for detecting a predisposition to development of PC, e.g. in prenatal tests. Detecting PCA3 protein allows differentiation between malignant and benign prostatic disease, and the level of PCA3 expression allows correlation with the grade of tumour. PCA3 protein and its fragments are also claimed to be useful in vaccines for preventing PC; in drug screens for identifying specific (antagonists (potentially useful therapeutically) and for studying protein-DNA interactions.

Sequence 3582 BP: 1052 A: 788 C: 679 G: 1063 T: 0 other:

Query Match	99.7%;	Score 2030.4;	DB 19;	Length 3582;
Best Local Similarity	99.6%;	Pred. No. 0;		
Matches 2028;	Conservative	5;	Mismatches	4; Indels 0; Gaps 0;
QY	1	agaagctggcatcagaaaaacagagggagagatttgtgtgctgcagccgagggagaccag	60	
Db	23	agaagctggcatcagaaaaacagagggagagatttgtgtgctgcagccgagggagaccag	82	
QY	61	gaagatctgcatggtgggaaggaacctgatgatacagaggaattacaacacataacttag	120	
Db	83	gaagatctgcatggtgggaaggaacctgatgatacagaggaattacaacacataacttag	142	
QY	121	tgtttcaatgaacaccaagaataaataagtgagagctagtcgcgtgtgagtcctcagt	180	
Db	143	tgtttcaatgaacaccaagaataaataagtgagagctagtcgcgtgtgagtcctcagt	202	
QY	181	gacacagggctggatcacccatcgacggcacctttctgagtactcagtcgagcaaaagaaga	240	
Db	203	gacacagggctggatcacccatcgacggcacctttctgagtactcagtcgagcaaaagaaga	262	
QY	241	ctacagacatctcaatggcaggggtgagaaataagaagaagctgtcgtactttaccatctga	300	
Db	263	ctacagacatctcaatggcaggggtgagaaataagaagaagctgtcgtactttaccatctga	322	
QY	301	ggccacacatctgtgaaatggagataattaacatcacatagaacacagcaagatgacaata	360	
Db	323	ggccacacatctgtgaaatggagataattaacatcacatagaacacagcaagatgacaata	382	
QY	361	taatgtctaagttagtgacatgtttttgcacatttccagccctttaaatatccacacaca	420	
Db	383	taatgtctaagttagtgacatgtttttgcacatttccagccctttaaatatccacacaca	442	
QY	421	caggagaacacaaagggaagcacagagatccctgggagaaaatgcccggcgcacatttggg	480	
Db	443	caggagaacacaaagggaagcacagagatccctgggagaaaatgcccggcgcacatttggg	502	
QY	481	tcatctgatgccctcgccctgtgctggctccgcttgtgaggggaagcacattagaaaatg	540	
Db	503	tcatctgatgccctcgccctgtgctggctccgcttgtgaggggaagcacattagaaaatg	562	
QY	541	aattgatgtgttccttaagagatgggcaggaaaaacagatcctgttgtgatatattttg	600	
Db	563	aattgatgtgttccttaagagatgggcaggaaaaacagatcctgttgtgatatattttg	622	
QY	601	aacgggattacagatttgaaatgaagtcaaaagtgcagattaccaatgcagaggaagaaca	660	
Db	623	aacgggattacagatttgaaatgaagtcaaaagtgcagattaccaatgcagaggaagaaca	682	
QY	661	gacgagaaaattctgtatggcttcaagacatgcaacaaacaaaatgaaatcagtgtgat	720	

New nucleic acid encoding prostate cancer antigen 3 - for diagnosis, prevention and treatment of prostatic cancer

Db 683 gacgagaaaatcttgatggtcttcacagacatgcaacaaacaaatggaatactgtgat 742
QY 721 acatgaggcagcaagctgggagagagataaccacagggcgagaggtcaggattctggcc 780
Db 743 acatgaggcagcaagctgggagagagataaccacagggcgagaggtcaggattctggcc 802
QY 781 ctgctgctaaactgtgctgtataaccacaaatcatttcatttataaccctcaaaacaa 840
Db 803 ctgctgctaaactgtgctgtataaccacaaatcatttcatttataaccctcaaaacaa 862
QY 841 agctgtgttaatactgatctctacaggttctctctgggcccacattctccatatccaa 900
Db 863 agctgtgttaatactgatctctacaggttctctctgggcccacattctccatatccaa 922
QY 901 gccacactcatttttaatttttagtctccagatctgactgtgaccttctcactgtag 960
Db 923 gccacactcatttttaatttttagtctccagatctgactgtgaccttctcactgtag 982
QY 961 aatacattactactttgttcaaaagaccctctggtgtgctgctcctaatatgtgactgact 1020
Db 983 aatacattactactttgttcaaaagaccctctggtgtgctgctcctaatatgtgactgact 1042
QY 1021 gtttttccaaaggagtgcttctggccaggggagctctgtgaacaggtgggaagcattctcaa 1080
Db 1043 gtttttccaaaggagtgcttctggccaggggagctctgtgaacaggtgggaagcattctcaa 1102
QY 1081 gatcttccaggggttactactactagcacacagcatgatactacagagtggaattatcta 1140
Db 1103 gatcttccaggggttactactactagcacacagcatgatactacagagtggaattatcta 1162
QY 1141 atcaacatcatcctcagtgcttggccatactgaaattcatttcccaactttgtgcccac 1200
Db 1163 atcaacatcatcctcagtgcttggccatactgaaattcatttcccaactttgtgcccac 1222
QY 1201 ttctcaagaccctcaaaatgctcattcattcaataatcacagattaaactttttttaacc 1260
Db 1223 ttctcaagaccctcaaaatgctcattcattcaataatcacagattaaactttttttaacc 1282
QY 1261 tggagaatcctaattgtacatgcagctatgggaatttaattacataattttgttccagt 1320
Db 1283 tggagaatcctaattgtacatgcagctatgggaatttaattacataattttgttccagt 1342
QY 1321 gcaagatgactaaagtcctttatccctccctcttgggtgattttttccagataaaagt 1380
Db 1343 gcaagatgactaaagtcctttatccctccctcttgggtgattttttccagataaaagt 1402
QY 1381 taaaatgcttagcctgtactaggtgtatagcacagcagcctcccccattccctccagc 1440
Db 1403 taaaatgcttagcctgtactaggtgtatagcacagcagcctcccccattccctccagc 1462
QY 1441 ctatctgtcatcaccatcaacccctcccatnysacctaataacaaatcacttgaatt 1500
Db 1463 ctatctgtcatcaccatcaacccctcccatnysacctaataacaaatcacttgaatt 1522
QY 1501 ccttgaaacatgtagcagacatacattttctcttctgctgagagctcttctctctctt 1560
Db 1523 ccttgaaacatgtagcagacatacattttctcttctgctgagagctcttctctctctt 1582
QY 1561 aantctagaatgataaagtttgaataagttgactatcttacttcatcagaagaagg 1620
Db 1583 aantctagaatgataaagtttgaataagttgactatcttacttcatcagaagaagg 1642
QY 1621 acacatagattcatcatcatcagacagcaataactaaagtgaattgtattata 1680
Db 1643 acacatagattcatcatcatcagacagcaataactaaagtgaattgtattata 1702
QY 1681 agagtttagataataatgaatgaagagkccacagagggaatttttatggggcacgtt 1740
Db 1703 agagtttagataataatgaatgaagagccacagagggaatttttatggggcacgtt 1762
QY 1741 tgaagcctgggagtggaamaagggcaggaaacctatgatttataataactt 1800
Db 1763 tgaagcctgggagtggaagcaagggggaacctatgatttataataactt 1822

QY 1801 catttctctatctctatccaaatccaaagctttttcacagaattcatgcagtcaaa 1860
Db 1823 catttctctatctctatccaaatccaaagctttttcacagaattcatgcagtcaaa 1882
QY 1861 tccccaaaaggtaacctttatccatttcattggtgagtgcgcttttagaattttggcaaatca 1920
Db 1883 tccccaaaaggtaacctttatccatttcattggtgagtgcgcttttagaattttggcaaatca 1942
QY 1921 tactgtcacttatctcacaactttgagatgtgttctctctgttagttaaattgaagaata 1980
Db 1943 tactgtcacttatctcacaactttgagatgtgttctctctgttagttaaattgaagaata 2002
QY 1981 gggcactctgtgagcacttttaggttcaactctgctcctggaataaagaatttacaaga 2037
Db 2003 gggcactctgtgagcacttttaggttcaactctgctcctggaataaagaatttacaaga 2059

RESULT 3
AAV62428
ID AAV62428 standard; cDNA; 1872 BP.
XX
AC AAV62428;
XX
DT 30-DEC-1998 (first entry)
XX
DE Prostate cancer antigen (PCA3) cDNA splice variant 2.
XX
KW Prostate cancer antigen cDNA splice variant 2; PCA3; prostatic cancer;
PC; ds.
XX
OS Homo sapiens.
XX
PN W09845420-AL.
PD 15-OCT-1998.
PF 09-APR-1998; 98WO-CA00346.
PR 10-APR-1997; 97US-0041836.
XX
PA (DIAG-) DIAGNOCURE INC.
PI Bussemakers MJG;
XX
DR WPI; 1998-568347/48.
XX
PT New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
prevention and treatment of prostatic cancer
XX
PS Claim 4; Pages 76-77; 11lpp; English.
XX
CC The present sequence represents the prostate cancer antigen (PCA3)
cDNA splice variant 2 sequence comprising of exons 1, 3, 4a and
4b of the PCA3 gene. The PCA3 cDNA splice variant 2 sequence,
isolated from a human primary prostatic tumour tissue cDNA library,
was found in approximately 65% of the cDNA clones isolated. The
invention claims for PCA3 cDNA variants and the proteins they encode.
The invention also claims for antibodies against PCA3 protein. The
antibodies are claimed to be useful for detecting PCA3 protein in
immunoassay tests, for diagnosing, assessing and prognosing of
prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin
or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed
to be useful for treating PC, while determining elevated levels of
PCA3 (as RNA or protein) is useful for detecting a predisposition
to development of PC, e.g. in prenatal tests. Detecting PCA3 protein
allows differentiation between malignant and benign prostatic disease,
and the level of PCA3 expression allows correlation with the grade of
tumour. PCA3 protein and its fragments are also claimed to be useful
in vaccines for preventing PC; in drug screens for identifying
specific (antagonists (potentially useful therapeutically) and for
studying protein-DNA interactions.
XX

SQ Sequence 1872 BP; 567 A; 389 C; 369 G; 539 T; 8 other;

Query Match 87.0%; Score 1772.4; DB 19; Length 1872;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1778; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 259 cagggtggaataagaagctgctgacttaccatctgagccacacatctgctgaa 318
|||
DB 94 cagaagtgaataagaagctgctgacttaccatctgagccacacatctgctgaa 153
|||
QY 319 atggagataattacatcactagaacagcaagatgacaaataatgtctaagtgtgac 378
|||
DB 154 atggagataattacatcactagaacagcaagatgacaaataatgtctaagtgtgac 213
|||
QY 379 atgtttttgcacatttcagcccttttaataatccacacacaggaagcacaagaa 438
|||
DB 214 atgtttttgcacatttcagcccttttaataatccacacacaggaagcacaagaa 273
|||
QY 439 gcacagagatccctgggagaaatgccggccgccatcttgggtcatcgatgagctcgcc 498
|||
DB 274 gcacagagatccctgggagaaatgccggccgccatcttgggtcatcgatgagctcgcc 333
|||
QY 499 ctgtgctcgtgctcgtgtgagggaagacattagaaaatgaattgatgtgttccttaa 558
|||
DB 334 ctgtgctcgtgctcgtgtgagggaagacattagaaaatgaattgatgtgttccttaa 393
|||
QY 559 aggatggcaggaataacagatcctgtgtggatatttttgaacgggattacagatttg 618
|||
DB 394 aggatggcaggaataacagatcctgtgtggatatttttgaacgggattacagatttg 453
|||
QY 619 aaatgaagtccaaagtgcattaccatgagagaaacagacagagaaatcttctgatg 678
|||
DB 454 aaatgaagtccaaagtgcattaccatgagagaaacagacagagaaatcttctgatg 513
|||
QY 679 gtttcaacagacatgcaacaaacaaatggaatactgtgatgacatgagcgcaagct 738
|||
DB 514 gtttcaacagacatgcaacaaacaaatggaatactgtgatgacatgagcgcaagct 573
|||
QY 739 ggggaggagataaccacgggcagagggtcaggattctgtgcccctgctgctaactgtgc 798
|||
DB 574 ggggaggagataaccacgggcagagggtcaggattctgtgcccctgctgctaactgtgc 633
|||
QY 799 gttcaaaccaaatcatttcataatttcaacccctcaaaacaaagctgtgttaatactga 858
|||
DB 634 gttcaaaccaaatcatttcataatttcaacccctcaaaacaaagctgtgttaatactga 693
|||
QY 859 tctctacggttccctctgtggcccaacattctccatatataccagccacactctttttaa 918
|||
DB 694 tctctacggttccctctgtggcccaacattctccatatataccagccacactctttttaa 753
|||
QY 919 attttagttccagatctgactgtgacattctctacactgataaataacattactcttt 978
|||
DB 754 attttagttccagatctgactgtgacattctctacactgataaataacattactcttt 813
|||
QY 979 gttcaaaagccctctggtgtgctgcctaataatgtagctgactgtttttcctaaggagt 1038
|||
DB 814 gttcaaaagccctctggtgtgctgcctaataatgtagctgactgtttttcctaaggagt 873
|||
QY 1039 tctgccccaggggatctgtgaacaggctgggaagcatctcaagatctttccagggttata 1098
|||
DB 874 tctgccccaggggatctgtgaacaggctgggaagcatctcaagatctttccagggttata 933
|||
QY 1099 ctactagcacacagcatgatcattacagagtggaattatctaatcaacatcctcagt 1158
|||
DB 934 ctactagcacacagcatgatcattacagagtggaattatctaatcaacatcctcagt 993
|||
QY 1159 gtctttgcccatactgaaatttcatttccacatttggcccattctcaagacctcaaaat 1218
|||
DB 994 gtctttgcccatactgaaatttcatttccacatttggcccattctcaagacctcaaaat 1053
|||
QY 1219 gtoattccataataatcacagattacatttttttttaacctgggaagaattcaatgtta 1278
|||||

DB 1054 gtcattccattaatatcacaggattaaacttttttttaacctgggaagaattcaatgtta 1113
QY 1279 catgcagctatgggaatttaattacatatattttgtttccagtgcaaaagactgaagtcc 1338
|||||
DB 1114 catgcagctatgggaatttaattacatatattttgtttccagtgcaaaagactgaagtcc 1173
|||||
QY 1339 ttatatccctccctctgtttgtatttttttccagataaaagttaaaatgcttagccttgt 1398
|||||
DB 1174 ttatatccctccctctgtttgtatttttttccagataaaagttaaaatgcttagccttgt 1233
|||||
QY 1399 actgaggctgtatacagacagcctctcccatccctccagccttatctgtcatcacat 1458
|||||
DB 1234 actgaggctgtatacagacagcctctcccatccctccagccttatctgtcatcacat 1293
|||||
QY 1459 caacccctcccatnysacctaaacaaaatctaactgttaattccttgaacatgtcagcnc 1518
|||||
DB 1294 caacccctcccatnysacctaaacaaaatctaactgttaattccttgaacatgtcagcnc 1353
|||||
QY 1519 atacattttcctctgctgagaagctctctctgtctcttaantctagaatgatgtaa 1578
|||||
DB 1354 atacattttcctctgctgagaagctctctctgtctcttaantctagaatgatgtaa 1413
|||||
QY 1579 agttttgaataagttgactatcttacttcatacgaagaaggacacatatgagattcatc 1638
|||||
DB 1414 agttttgaataagttgactatcttacttcatacgaagaaggacacatatgagattcatc 1473
|||||
QY 1639 atcacatgagacagcaaaatactaaaagtgaatttgattataagagtttagataaata 1698
|||||
DB 1474 atcacatgagacagcaaaatactaaaagtgaatttgattataagagtttagataaata 1533
|||||
QY 1699 tgaatatgaagkccacacagaggaatgtttatggggcacgtttgtaagcctgggagtga 1758
|||||
DB 1534 tgaatatgaagkccacacagaggaatgtttatggggcacgtttgtaagcctgggagtga 1593
|||||
QY 1759 agmaaggcaggaacctcatagttatttataataataacttctctctctctctctc 1818
|||||
DB 1594 agmaaggcaggaacctcatagttatttataataataacttctctctctctctctc 1653
|||||
QY 1819 acaatatccaaacagcttttcacagaattcatgcagtgcaaatccccaaagtaaccttt 1878
|||||
DB 1654 acaatatccaaacagcttttcacagaattcatgcagtgcaaatccccaaagtaaccttt 1713
|||||
QY 1879 atccatttcatggtgagtcgcttttagaattttggcaaatcatctggtcacttatctca 1938
|||||
DB 1714 atccatttcatggtgagtcgcttttagaattttggcaaatcatctggtcacttatctca 1773
|||||
QY 1939 actttgagatgtgtttgtccttctgtagtttaattgaaagaataaggcactctgtgagcca 1998
|||||
DB 1774 actttgagatgtgtttgtccttctgtagtttaattgaaagaataaggcactctgtgagcca 1833
|||||
QY 1999 ctttaggggttcaactcctgcataaagaatttacaaga 2037
|||||
DB 1834 ctttaggggttcaactcctgcataaagaatttacaaga 1872
|||||

RESULT 4
AAA06687
ID AAA06687 standard; cDNA; 3112 BP.
XX
AC AAA06687;
XX
DT 13-JUN-2000 (first entry)
XX
DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:468.
XX
KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW immunogenic; cytostatic; vaccine; ss.
OS Homo sapiens.
XX
PN WO200004149-A2.
XX
PD 27-JAN-2000.

XX 14-JUL-1999; 99WO-US15838.
PF
XX 14-JUL-1998; 98US-0115453.
PR 14-JUL-1998; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX
XX (CORI-) CORIXA CORP.
XX
XX Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;
PI WPI; 2000-171268/15.
XX
XX New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumor protein -
PT
XX
XX Claim 1; Page 259-260; 263pp; English.
XX
XX The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate antibodies
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the polynucleotides encoding the polypeptides can be used as a probe or
CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC AAY82000 to AAY82020 represent sequences used in the exemplification of
CC the present invention.
XX
XX Sequence 3112 BP; 975 A; 587 C; 624 G; 926 T; 0 other;
SQ

Query Match 85.3%; Score 1737.8; DB 21; Length 3112;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 1772; Conservative 4; Mismatches 8; Indels 3; Gaps 3;

QY 252 tcaatggcagggggtgagaataagaaggctgctgactttaccatctgagggccacacatc 311
DB 1302 tcaactaaataggtgagaataagaaggctgctgactttaccatctgagggccacacatc 1361
QY 312 tgcgtaaatggagataattaacatcactagaacacagacagatgacaataatgtctaag 371
DB 1362 tgcgtaaatggagataattaacatcactagaacacagacagatgacaataatgtctaag 1421
QY 372 tagtgacatgtttttgcacatttccagcccttttaataatccacacacagagaagcaca 431
DB 1422 tagtgacatgtttttgcacatttccagcccttttaataatccacacacagagaagcaca 1481
QY 432 aaagggaacacagagatccctgggagaataatgccggccgccattcttgggtcatcgatgag 491
DB 1482 aaagggaacacagagatccctgggagaataatgccggccgccattcttgggtcatcgatgag 1541
QY 492 cctcgcccttgcctggtcccgcttggagggaagacattagaaatgaattgatgtgt 551
DB 1542 cctcgcccttgcctggtcccgcttggagggaagacattagaaatgaattgatgtgt 1601
QY 552 tccttaaaagatgggcaggaaacagatcctgttggatattatttgaacggggattac 611
DB 1602 tccttaaaagatgggcaggaaacagatcctgttggatattatttgaacggggattac 1661
QY 612 agattgaaatgaagtccaaagtggacattaccatagagagaaacagacagagaaaat 671
DB 1662 agattgaaatgaagtccaaagtggacattaccatagagagaaacagacagagaaaat 1721
QY 672 ctgtgagcttcacagacatgacaacaaataatgaatactgtgatgacatgagggcag 731

DB 1722 ctgtgagcttcacagagacatgcaacaaacaaatggaatactgtgatgacatgagggcag 1781
QY 732 ccaagctggggaggagataaccacaggggagagggtcagagattcttgcctctgctgctaa 791
DB 1782 ccaagctggggaggagataaccacaggggagagggtcagagattcttgcctctgctgctaa 1841
QY 792 actgtgcgttcataaaccaaatcatttcatttctaacctcctcaaaacaaagctgtgttaa 851
DB 1842 actgtgcgttcataaaccaaatcatttcatttctaacctcctcaaaacaaagctgtgttaa 1901
QY 852 tatctgatctctacaggttctctctggggcccaacattctccatatatccagccacactcat 911
DB 1902 tatctgatctctacaggttctctctggggcccaacattctccatatatccagccacactcat 1961
QY 912 ttttaataatttagttccagatctgtactgtgacctttctacactgtagaataacattac 971
DB 1962 ttttaataatttagttccagatctgtactgtgacctttctacactgtagaataacattac 2021
QY 972 tcaattttgtcaaaagaccctctgtgtgctgctaaatagtactgtactgtttttctctaa 1031
DB 2022 tcaattttgtcaaaagaccctctgtgtgctgctaaatagtactgtactgtttttctctaa 2081
QY 1032 ggaagtcttgcggccagggtatctgtgaacagagctgggaagcatctcaagatctttccag 1091
DB 2082 ggaagtcttgcggccagggtatctgtgaacagagctgggaagcatctcaagatctttccag 2141
QY 1092 ggttactactactagcacacagcatgatcattacgagtgagtgaaattctaatcaacatcat 1151
DB 2142 ggttactactactagcacacagcatgatcattacgagtgagtgaaattctaatcaacatcat 2201
QY 1152 cctcagtgctttggcccatactgaaattcatttcccacactttgtgcccatttctcaagacc 1211
DB 2202 cctcagtgctttggcccatactgaaattcatttcccacactttgtgcccatttctcaagacc 2261
QY 1212 tcaaaatgcatctccattcaataatcacagagattaaacttttttttaacctgggaagaattc 1271
DB 2262 tcaaaatgcatctccattcaataatcacagagattaaacttttttttaacctgggaagaattc 2321
QY 1272 aatgttacatgcatgctatgggaatttaattacataatttgttttccagtgcagaagatgac 1331
DB 2322 aatgttacatgcatgctatgggaatttaattacataatttgttttccagtgcagaagatgac 2381
QY 1332 taagtctcttataccctcccttctgtgtgttttttccagtaataagtttaaatgtctta 1391
DB 2382 taagtctcttataccctcccttctgtgtgttttttccagtaataagtttaaatgtctta 2441
QY 1392 gccttgactgagctgtgtatacag-cacagcctctcccatccctccagcttattctgtc 1450
DB 2442 gccttgactgagctgtgtatacagcctctcccatccctccagcttattctgtc 2501
QY 1451 atcacatcaacccctcccatnysacataacaaatactaaactgttaattccttgaacat 1510
DB 2502 atcacatcaacccctcccatg-cacctaacaataactaaactgttaattccttgaacat 2560
QY 1511 gtcagncatatacttrttctctctgctgagaagctcttctgtctcttcttaantctagaa 1570
DB 2561 gtcagg-catacatattctctctgctgagaagctcttctgtctcttcttaantctagaa 2619
QY 1571 tgatgtaaagttttgaataagtgtactattctactcatgcaagaagggacacatatga 1630
DB 2620 tgatgtaaagttttgaataagtgtactattctactcatgcaagaagggacacatatga 2679
QY 1631 gattcatcatcatgagacagcaataactaaaagtgtattgtattgtataagaggtttaga 1690
DB 2680 gattcatcatcatgagacagcaataactaaaagtgtattgtattgtataagaggtttaga 2739
QY 1691 taaatatagaatgcaagkccacagaggggaatgtttatggggcacgtttgttaagcctg 1750
DB 2740 taaatatagaatgcaagkccacagaggggaatgtttatggggcacgtttgttaagcctg 2799
QY 1751 ggaatgtgaagaaagcagggaaacctcatgtattcttataataatacttcttctta 1810
DB 2800 ggaatgtgaagaaagcagggaaacctcatgtattcttataataatacttcttctta 2859

QY 1811 tctctatcacatataccaagcttttcacagaaattcatgctgagtcagaaatccccaaagg 1870
|||||
Db 2860 tctctatcacatataccaagcttttcacagaaattcatgctgagtcagaaatccccaaagg 2919
QY 1871 taacctttatccatttcattgagtgagtgagcttttagaattttggcaaatcactactggtcac 1930
|||||
Db 2920 taacctttatccatttcattgagtgagtgagcttttagaattttggcaaatcactactggtcac 2979
QY 1931 ttatctcaactttgagatgtgttttctctgttagttaattgaaagaaatagggaactctt 1990
|||||
Db 2980 ttatctcaactttgagatgtgttttctctgttagttaattgaaagaaatagggaactctt 3039
QY 1991 gtgagccactttgaggttcactccctggcaataaagaatttcaaaaga 2037
|||||
Db 3040 gtgagccactttgaggttcactccctggcaataaagaatttcaaaaga 3086

RESULT 5
AAA06689/c
ID AAA06689 standard; cDNA; 2426 BP.
XX
AC AAA06689;
XX
DT 13-JUN-2000 (first entry)
XX
DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:470.
XX
KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW immunogenic; cytostatic; vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200004149-A2.
XX
PD 27-JAN-2000.
XX
PF 14-JUL-1999; 99WO-US15838.
XX
PR 14-JUL-1998; 98US-0115453.
PR 14-JUL-1998; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX
PA (CORI-) CORIXA CORP.
XX
PI Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;
XX
DR WPI; 2000-171268/15.
XX
PT New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumor protein -
XX
PS Claim 1; Page 261-262; 263pp; English.
XX
CC The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (ptp). The polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate antibodies
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the polynucleotides encoding the polypeptides can be used as a probe or
CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC AAA82000 to AAA82020 represent sequences used in the exemplification of
XX the present invention.
SQ Sequence 2426 BP; 717 A; 476 C; 548 G; 685 T; 0 other;

Query Match 85.0%; Score 1731.4; DB 21; Length 2426;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1762; Conservative 4; Mismatches 2; Indels 3; Gaps 3;
QY 263 ggtgagaataagaagcgctgctgactttaccatctcgagggccacacatctgctgaaatgg 322
|||||
Db 1769 GGTGAGAAATAGAAAGGCTGCTGACTTTACCATCTGAGGCCACACATCTGCTGAATGG 1710
QY 323 agataataatcaatcactactagaacagagatgacaataataatgtctaaagttagtgcacatgt 382
|||||
Db 1709 AGATATTAACATCACTAGAAAACAGACAGATGACATAATAATGTCTAAGTAGTGCATGT 1650
QY 383 ttttgacatttcagcccttttaaatatccacacacagagaagcaaaaggaagcac 442
|||||
Db 1649 TTTTGACATTTCCAGCCCTTTAAATATCCACACACAGGAAGCAGAAAAGGAAGCAC 1590
QY 443 agagatccctgggagaaaatgcccggccgcatcttgggttcacatcgatgagcctcgccctgt 502
|||||
Db 1589 AGAGATCCTCGGAGAAATGCCCGCGCCCATCTTTGGGTTCATCGATGAGCCTCGCCCTGT 1530
QY 503 gctgtgtccgcctgtgaggggaagacattagaaaaatgaattgattgtctcccttaaaaga 562
|||||
Db 1529 GCTGTGTCGCCCTTGTGAGGGAAGGACATTAGAAAATGAATGTGTTCTCTTAAAGGA 1470
QY 563 tggcgagaaaacagatcctgtgtggatatttttgaacgggattacagatttgaat 622
|||||
Db 1469 TGGGAGGAGAAAACAGATCCTGTGTGGATATTTATTGAAACGGGATTTACAGATTTGAAAT 1410
QY 623 gaagtcaaaaagtgcattaccaatgagaggaaaacacagacagaaaaatcttgatggctt 682
|||||
Db 1409 GAAGTCACAAAGTGAGCATTTACCAATGAGGGAAGCAACACGAGAAAATCTTGATGGCTT 1350
QY 683 cacaagacatgcaacaacaaaatggaaatactgtgatgacatgagggagccaagctgggg 742
|||||
Db 1349 CACAAGACATGCAACAAACAAAATGGAATACTGTGATGACATGAGCAGCAAGCTGGGG 1290
QY 743 aggaataaacacggggcagaggggtcaggattctggccctgctgctaaactgtcgcttc 802
|||||
Db 1289 AGGATAAACCCAGCGGCGAGGGGTACAGGATTTCTGGCCCTGCTGCTAAACTGTGCGTTTC 1230
QY 803 ataaccaaaatctttcatttctaacctcacaacaaagctgtgtgtaataatctgacatcc 862
|||||
Db 1229 ATAACCAAAATCATTTTCATATTTCTAACCCCTCAAAACAAAGCTGTGTATATCTGATCTC 1170
QY 863 tacggttcctctgtgggcccacacattctccatataatccagccacactcaatttttaatttt 922
|||||
Db 1169 TACGGTTCTCTGGGCCCAACATTTCTCCATATATCCAGCCACACTCATTTTAAATATTT 1110
QY 923 agttccagatctgactgtgacctttctacactgtagaataacattactattttgttc 982
|||||
Db 1109 AGTTCACAGATCTGACTGTGACCTTTCTACACTGTAGAAATACATTAATCTCAATTTTGTTC 1050
QY 983 aaagaccctctgtgtgctgctaataatgactgactgtttttccaaaggaggttcttg 1042
|||||
Db 1049 AAAGACCCTTCGTGTGCTGCCTAATATGATGACTGACTGTTTCTCTAAGGAGTGTCTG 990
QY 1043 gccaggggatctgtgaacaggtcggaagcatctccaaagattttccagggttatactta 1102
Db 989 GCCCAGGGGATCTGTGAACAGGCTGGGAAGCATCTCAAGATCTTCCAGGGTTATACTTFA 930
QY 1103 ctgacacacagatcatcattacggagtgaaattatctaataacatcatcctcagtgct 1162
Db 929 CTAGCACACAGCATGATCATTTACGGAGTGAATATATCTAATCAACATCATCTCCAGTGCTCT 870
QY 1163 ttggccactactgaaatttcacatttccacatttgtgccatttccaaagacctcaaaatgtca 1222
Db 869 TTGCCCATACTGAATTCATTTCCCACTTTTGTGCCCACTTCTCAAGACCTCAAAATGTCA 810
QY 1223 ttccattaatatcacaggattaaacttttttttaacctgggaagaattcaattgttacatg 1282
Db 809 TTCCATTAAATATCACAGGATTAACTTTTTTTTTTTTAACTTGAAGAAATTCATTTTACATG 750

QY 1283 cagctatgggaatttaattacatatatttcttccagtgcaaaagatgactaaagtccttta 1342
|||||
Db 749 CAGCTATGGGAATTAATACATATTTGTTTCCAGTGCAAAAGATGACTAAGTCCTTTA 690
QY 1343 tccctccctttgttgatttttttccagtgataaaagttaaaatgcttgacctgtactg 1402
|||||
Db 689 TCCTCCCTCTTGTGATTTTTTTTCCAGTATAAAAGTTAAATGCTTAGCCTTGTACTG 630
QY 1403 aggtgtatcacg-cacagcctctcccatccctccagccttatctgtcatcacatcaa 1461
|||||
Db 629 AGGCTGTATACAGCCACAGCCTCTCCCATCTCCAGCCTTAICTGTATCATCAACATCA 570
QY 1462 cccctcccatnysaccatacaaaaataactgttaattccctgaacatgtcaggncata 1521
||||| :|||||
Db 569 CCCTCCCATG-CACCTAAACAAATCTAACTTGTAATCTTGAACATGTCAGG-CATA 512
QY 1522 catttccctctgctggaagccttccctgtctcttaantctcagaatgatgtaagt 1581
|||||
Db 511 CAPTATTCCTCTGCTGAGAGCTCTCCTTGCTCTTAAATCTAGAATGATGTAAGT 452
QY 1582 ttgtaaatgtagactatcttcatcacaagaaggacacacatatagattcatcatc 1641
|||||
Db 451 TTGTAATAAGTTGACTATCTTACTTCATGCAAAAGAGGACACATATGAGATTCATCATC 392
QY 1642 acatgagacagcaaatactaaaaagtgtaatttgattataagagtttagataaataatga 1701
|||||
Db 391 ACATGAGACAGCAAAATACTAAAAGTGTAATTTGATTATAAGAGTTTAGATAAATATATGA 332
QY 1702 aatgcaagakccacagaggaatgtttatggggcacgtttgtaagcctgggatgtaagm 1761
|||||
Db 331 AATGCAAGAGCCACAGAGGGAATGTTATGTTGAGGACGTTTGTAAAGCCTGGGATGTAAGC 272
QY 1762 aaaggcaggaacccatagatcttatataataatcattcattctctctctcatcaca 1821
|||||
Db 271 AAGGCGAGGAACCTCATGATATCTTATATATAATATATCTTCTATCTCTATCATCA 212
QY 1822 atatcaacaagcttttcagaaattcagtgagtgcaaatcccaaaagtgtaacctttatc 1881
|||||
Db 211 ATATCCAACAAGCTTTTCAGAAATTCATGAGTGCAATCCCAAAAGGTAACTTTATC 152
QY 1882 catttcatagtgagtgcgctttagaatttttgcaaatcactactggtcacttatctcaact 1941
|||||
Db 151 CAPTTATGGTGAGTGGCGCTTTAGAAATTTGGCAAAATCATACTGGTCACTTATCTCAACT 92
QY 1942 ttgagatgtgttctgtctgttgatttaattgaaagaataaggcactcttctgtgagccact 2001
|||||
Db 91 TTGAGATGTGTTGTCTTGTAGTTAATTGAAGAATAATAGGCACCTCTTGTGAGGCCACTT 32
QY 2002 taggggtcactcctggcaataaagaatttac 2032
|||||
Db 31 TAGGGTTCACTCTCGGCAATAAAGAATTAC 1

RESULT 6
AAA06688/c
ID AAA06688 standard; cdna; 2229 BP.
XX
AC AAA06688;
XX
XT 13-JUN-2000 (first entry)
XX
DT Human immunogenic prostate tumour protein cdna sequence SEQ ID NO:469.
DE Human immunogenic prostate tumour protein cdna sequence SEQ ID NO:469.
XX
KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW immunogenic; cytostatic; vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200004149-A2.
XX
PD 27-JAN-2000.
XX

PF 14-JUL-1999; 99WO-US15838.
XX
PR 14-JUL-1998; 98US-0115453.
PR 14-JUL-1998; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX (CORI-) CORIXA CORP.
PA
PI Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;
PI WPI; 2000-171268/15.
DR
XX
XX
PT New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumor protein -
XX
PS Claim 1; Page 260-261; 263pp; English.
XX
CC The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (Prp). The polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate antibodies
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the polynucleotides encoding the polypeptides can be used as a probe or
CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC AAY82000 to AAY82020 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 2229 BP; 654 A; 447 C; 481 G; 647 T; 0 other;
Query Match 84.1%; Score 1712.4; DB 21; Length 2229;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1765; Conservative 4; Mismatches 2; Indels 5; Gaps 5;
QY 263 ggtgagaaataagaaggctgtgactttaccatctgagggccacacatctgtgaaatgg 322
|||||
Db 1775 GGTGAGAAATAAGAAGGCTGCTGTACTTTACCATCTGAGGCCACACATCTGCTGAAATGG 1716
QY 323 aqataaataacatcactagaacaagcaaatgacaataataatgcttaagttagcatgt 382
|||||
Db 1715 AGATAAATTAACATCATAGAAAACAGCAAGATGACAAATATAATGTCTAAAGTAGTGACATGT 1656
QY 383 ttttgacacatttccagccctttaaataatccacacacaggaagcacacaaaaggagcac 442
|||||
Db 1655 TTTTGCACATTTCCAGCCCTTTAAATATCCACACACAGGAGGACCAAAAGGAGACAC 1596
QY 443 agagatccctgggagaaatgcccggccgcatcttgggttcctcgatgagcctcgccctgt 502
|||||
Db 1595 AGAGATCCCTGGGAGAAATGCCCGCCGCATCTTGGGTGTCATGATGAGCCTCGCCCTGT 1536
QY 503 gcttggtcccgctgtgagggaggaagacattagaaaaatgaattgattgttcccttaagga 562
|||||
Db 1535 GCCTGTCCTCCGCTTGTGAGGGAAGGACATTAGAAAATGAATGTGATGTGTTCTTAAAGGA 1476
QY 563 tgggcaggaaaaacagatcctgtgtggatattatttgaacgggattacagatttgaat 622
|||||
Db 1475 TGGGCAGGAAAAACAGATCCTGTTGTGGATATTTATTTGAACGGGATTACAGATTGCAAT 1416
QY 623 gaagtcaaaaagtgtgagcattaccgaatgagggaaacacagacgagaaaaattctgtgacct 682
|||||
Db 1415 GAAGTCACAAAAGTGAGCATTTACCAATGAGAGGAAAAACAGACGAGAAAAATCTTGTGCTT 1356
QY 683 cacaagacatgcaacaaacaaaataactgtgactgtgacatgagcagccagctgggg 742
|||||
Db 1355 CACAAGACATGCAACAAACAAAATGGAATGATGTGTGATGATGATGAGGACGCCAAGCTGGGG 1296

Qy	1822	atatccacaagcttttccacagaatctatcagtcagtgcaaatccccaaaggttaacctttatc	1881
Db	219	ATATCCCAACAAAGCTTTTTCACAGAAATTCATGCACTGCAAAATCCCAAGGTAACTTTATC	160
Qy	1882	catttcattgtagtcgcttttagaattttgcaaatcatactgctcaacttctcaact	1941
Db	159	CATTTCATATGTTGAGTGGCGCTTTTGAATTTTGGCAAATCATACTGGTGCATTTATCTCAACT	100
Qy	1942	ttgagatgtgttttgcctctttagttaaattgaagaataataggcactcttcttgagccactt	2001
Db	99	TTGAGATGTGTTTGTCTCTACTTAATTCAGAAATATAGGCACTCTTGTGAGCCACTT	40
Qy	2002	taggttccatcctctggccaataagaattaccaaaaga	2037
Db	39	TAGGTTTCACTCCTGGCAATAAAGAATTTACAAAGA	4
RESULT	7		
AAV62429			
ID	AAV62429	standard; cDNA; 820 BP.	
XX	AC	AAV62429;	
XX	30-DEC-1998	(first entry)	
XX	Prostate cancer antigen (PCA3)	cDNA splice variant 3.	
XX	Prostate cancer antigen cDNA splice variant 3; PCA3;	prostatic cancer;	
KW	PC; ds.		
XX	Homo sapiens.		
XX	WO9845420-A1.		
PN	15-OCT-1998.		
PD			
XX	09-APR-1998;	98WO-CA00346.	
XX	10-APR-1997;	97US-0041836.	
XX	(DIAG-) DIAGNOCURE INC.		
XX	Bussemakers MJG;		
PI	WPI;	1998-568347/48.	
DR			
XX	New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,		
PT	prevention and treatment of prostatic cancer		
XX	Claim 4; Pages 77-78; 111pp; English.		
XX	The present sequence represents the prostate cancer antigen (PCA3)		
CC	cDNA splice variant 3 sequence comprising of exons 1, 3, and 4a		
CC	of the PCA3 gene. The PCA3 cDNA splice variant 3 sequence,		
CC	isolated from a human primary prostatic tumour tissue cDNA library,		
CC	was found in approximately 15% of the cDNA clones isolated. The		
CC	invention claims for PCA3 cDNA variants and the proteins they encode.		
CC	The invention also claims for antibodies against PCA3 protein. The		
CC	antibodies are claimed to be useful for detecting PCA3 protein in		
CC	immunoassay tests, for diagnosing, assessing and prognosing of		
CC	prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin		
CC	or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed		
CC	to be useful for treating PC, while determining elevated levels of		
CC	PCA3 (as RNA or protein) is useful for detecting a predisposition		
CC	to development of PC, e.g. in prenatal tests. Detecting PCA3 protein		
CC	allows differentiation between malignant and benign prostatic disease		
CC	and the level of PCA3 expression allows correlation with the grade of		
CC	tumour. PCA3 protein and its fragments are also claimed to be useful		
CC	in vaccines for preventing PC; in drug screens for identifying		
CC	specific (ant)agonists (potentially useful therapeutically) and for		
CC	studying protein-DNA interactions.		
XX	Sequence 820 BP; 262 A; 169 C; 191 G; 198 T; 0 other;		
SQ			

XX	PN	WO200004149-A2.
XX	XX	
XX	PD	27-JAN-2000.
XX	XX	
XX	PF	14-JUL-1999; 99WO-US15838.
XX	XX	
XX	PR	14-JUL-1998; 98US-0115453.
PR	14-JUL-1998; 98US-0116134.	
PR	23-SEP-1998; 98US-0159812.	
PR	23-SEP-1998; 98US-0159822.	
PR	15-JAN-1999; 99US-0232149.	
PR	15-JAN-1999; 99US-0232880.	
PR	09-APR-1999; 99US-0288946.	
XX	XX	(CORI-) CORIXA CORP.
PA	XX	
PI	XX	Dillon DC, Harlocker SL, Yugu J, Xu J, Mitcham JL;
XX	XX	
DR	XX	WPI; 2000-171268/15.
XX	XX	
PT	XX	New polypeptide useful for treating and diagnosing prostate cancer
PT	XX	comprises an immunogenic portion of prostate tumor protein -
XX	XX	
PS	XX	Claim 1; Page 262; 263pp; English.
XX	XX	
CC	XX	The present invention describes isolated polypeptides, comprising an
CC	XX	immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC	XX	and polynucleotides encoding them have cytostatic activity and can be
CC	XX	used in vaccines and in gene therapy. The polypeptides and
CC	XX	polynucleotides encoding them, antigen presenting cells which express
CC	XX	the polypeptides, antibodies against the polypeptides and vaccines
CC	XX	comprising them can be used for inhibiting the development of prostate
CC	XX	cancer in a patient. The polypeptides can be used to generate antibodies
CC	XX	or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC	XX	the polynucleotides encoding the polypeptides can be used as a probe or
CC	XX	to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC	XX	AA082000 to AA082020 represent sequences used in the exemplification of
CC	XX	the present invention.
XX	XX	
SQ	XX	Sequence 812 BP; 198 A; 189 C; 169 G; 256 T; 0 other;
Query Match 35.5%; Score 722.4; DB 21; Length 812;		
Best Local Similarity 99.9%; Pred. No. 9.7e-198;		
Matches 723; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
QY	259	cagggtgagaaataagaaggctgctgactttaccatctgagggccacacatctgctgaa 318
DB	724	CAGAGGTGAGAATAAGAAAGGCTGCTGACTTTACCATCTGAGGCCACACATCTGCTGAA 665
QY	319	atggagataataaacatcactagaaacagcaagatgacataataatgtctaaagttagtgac 378
DB	664	ATGGAGATANTTAACATCACTAGAACACAGACAGATCAATATATGTTCTAGTAGTGAC 605
QY	379	atgttttgcacatttccagccctttaaataatccacacacaggaagacacaaaaggaa 438
DB	604	ATGTTTTTGCACATTTCCAGCCCTTTAAATATCCACACACAGGAAGCACAAAAGGAA 545
QY	439	gcacagagatccctgggagaaatgcccggcccatcttgggtcctcgtatgagcctcgcc 498
DB	544	GCACAGAGATCCTTGGGAGAAATGCCGCCCATCTTTGGGTCAATCGATGAGCCCTCGCC 485
QY	499	ctgtcctggtcccgcttggaggaagacattagaaaatgaattgattgtgttctctaa 558
DB	484	CTGTGCTGCTCCCGCTTGTGAGGGAGGACATAGAAAATGAATGATGTGTCTCTTAA 425
QY	559	aggatgggcaggaacacagatcctgtgtggtattttgaacgggattacagatttg 618
DB	424	AGGATGGGCAGGAACACAGATCCTGTGTGGATATTTATTGAAACGGGATTACAGATTG 365
QY	619	aaatgaagtcaaaaatgagcattaccatgagaggagaaacagacagagaaaatcttgatg 678

Db 364 AATAGAGTCACAAAGTGAGCATTACCAATGAGAGGAAACACAGACGAGAAATCTTGATG 305
Qy 679 gcttcacagacatgaacaaatggaatactgtgatgacatgagcagccaaagct 738
Db 304 GCTTCACAGACATGCACAAACAAATGGAATACTGTGATGACATGAGGAGCCCAAGCT 245
Qy 739 ggggaggagataaacacacggggcagagggtcagggtctggtccctgctgcctaaactgtgc 798
Db 244 GGGGAGGAGATAACACCGGGGAGAGGGTCAGGATCTGCGCCCTGCTGCTAAACTGTGC 185
Qy 799 gttcataacaaaatcatttcatttctaacccctcaaaacaaagctgtgttaatactga 858
Db 184 GTTTCATAACCAATCATTTTCATTTCTTAACCCCTCAAAACAAAGCTGTGTAATATCTGA 125
Qy 859 tctctacggttctctctggggcccaacattctccatatatccagccacacactcattttta 918
Db 124 TCTCTACGGTCTCTCTGGGCCACATCTCCATATATCCAGCCACACTCATTTTAAAT 65
Qy 919 atttagttccagatctgtactgtgacatttctacactgtagaataaacactactcatttt 978
Db 64 ATTTAGTTCCAGATCTGTACTGTGACCTTTCTACACTGTAGATAACATTACTCATTTT 5
Qy 979 gttc 982
Db 4 GTTC 1

RESULT 9
AA37486
ID AAX37486 standard; cDNA; 597 BP.
XX
AC AAX37486;
XX
DT 06-JUL-1999 (first entry)
XX
DE Human secreted protein cDNA fragment containing gene 36.
XX
KW Human; secreted protein; treatment; prevention; protein therapy; AIDS;
KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW immune system disease; autoimmune disease; hepatic disease; lymphoma;
KW renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
KW cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
KW pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
KW arthritis; malignancy; digestive; endocrine; infection; ss.
XX
OS Homo sapiens.
XX
PN W09918208-A1.
XX
PD 15-APR-1999.
XX
PF 01-OCT-1998; 98WO-US20775.
XX
PR 02-OCT-1997; 97US-0060884.
PR 02-OCT-1997; 97US-0060833.
PR 02-OCT-1997; 97US-0060836.
PR 02-OCT-1997; 97US-0060837.
PR 02-OCT-1997; 97US-0060838.
PR 02-OCT-1997; 97US-0060839.
PR 02-OCT-1997; 97US-0060843.
PR 02-OCT-1997; 97US-0060862.
PR 02-OCT-1997; 97US-0060866.
PR 02-OCT-1997; 97US-0060874.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Carter KC, Duan DR, Endress GA, Peng P, Ferrie AM;
PI Florence KA, Greene JM, Janat F, Lafleur DW, Ni J;
PI Rosen CA, Ruben SM, Shi Y, Young P, Yu G;
XX
DR WPI; 1999-264022/22.
DR P-PSDB; AAY07887.

XX New isolated human genes and the secreted polypeptides they encode
PT Claim la; Page 247; 368pp; English.
PS
XX
CC This invention describes novel isolated human genes and the secreted
CC proteins they encode. The products of the invention are useful for
CC preventing, treating or ameliorating medical conditions, e.g. by protein
CC or gene therapy. Also pathological conditions can be diagnosed by
CC determining the amount of the new polypeptides in a sample or by
CC determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 101 polynucleotides, based
CC on which tissues they are most highly expressed in, and include
CC developing products for the diagnosis or treatment of cancer, tumours,
CC neurodegenerative disorders, developmental abnormalities and fetal
CC deficiencies, blood disorders, leukemias, diseases of the immune system,
CC autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,
CC allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate
CC disease, skeletal or cardiac muscle disorders, pulmonary disorders,
CC transplant rejection, disorders involving osteoclasts such as
CC osteoporosis, arthritis or malignancies, digestive/endocrine disorders,
CC infections and AIDS. The human secreted proteins of the invention are
CC represented in AAY07852-Y07993 and the encoding nucleic acids are
CC represented in AAX37451-X37552.
XX
SQ Sequence 597 BP; 181 A; 131 C; 134 G; 150 T; 1 other;

Query Match 25.7%; Score 524; DB 20; Length 597;
Best Local Similarity 99.6%; Pred. NO. 9.9e-141;
Matches 524; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 462 gccggcgcccatcttggtcctgcatgagcctgcctgctgcctggtcctgctgtgag 521
Db 54 gccggcgcccatcttggtcctgcatgagcctgcctgctgcctggtcctgctgtgag 113
Qy 522 ggaagcattagaaatgaattgattgttctttaaaggatggcagaaacagatcc 581
Db 114 ggaagcattagaaatgaattgattgttctttaaaggatggcagaaacagatcc 173
Qy 582 tgttggtatatttttgaacgggattacagatttgaattgaatgaatgaatgaatgaat 641
Db 174 tgttggtatatttttgaacgggattacagatttgaattgaatgaatgaatgaatgaat 233
Qy 642 taccatgagagaaacacagacgaaatcttgccttcacagacatgcacaac 701
Db 234 taccatgagagaaacacagacgaaatcttgccttcacagacatgcacaac 293
Qy 702 aaatggaatactgtgatgacatgaggcagccaaagctgggaggagataaacacgggca 761
Db 294 aaatggaatactgtgatgacatgaggcagccaaagctgggaggagataaacacgggca 353
Qy 762 gaggtcaggattctgcccctgctgcctaaactgctgttcataaccataatttcata 821
Db 354 gaggtcaggattctgcccctgctgcctaaactgctgttcataaccataatttcata 413
Qy 822 ttcttaaccccaaaacaaagctgttgaatatctgactctacgttctctctgagccc 881
Db 414 ttcttaaccccaaaacaaagctgttgaatatctgactctacgttctctctgagccc 473
Qy 882 aacattctccatataccagccacactcatttttaattattagttccagatctgactg 941
Db 474 aacattctccatataccagccacactcatttttaattattagttccagatctgactg 533
Qy 942 tgacctttctacactgtagaataacattactactcattttgttcaaaa 987
Db 534 tgacctttctacactgtagaataacattactactcattttgttcaaaa 579

RESULT 10
AAA06545
ID AAA06545 standard; cDNA; 718 BP.
XX

Query Match 15.7%; Score 319.6; DB 21; Length 437;
Best Local Similarity 98.5%; Pred. No. 5,3e-82;
Matches 319; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 259 cagggtgagaaataagaagctgctgaactttacatctgagggccacacatctgctgaa 318
DB 114 cagaggtgagaaataagaagctgctgaactttacatctgagggccacacatctgctgaa 173
QY 319 atggagataaataacatcactagaacacagcaagatgacaataataatgtctaaagttagtac 378
DB 174 atggagataaataacatcactagaacacagcaagatgacaataataatgtctaaagttagtac 233
QY 379 atgtttttgacatttcagcccttttaataatccacacacagcaagcaaaaaggaa 438
DB 234 atgtttttgacatttcagcccttttaataatccacacacagcaagcaaaaaggaa 293
QY 439 gcacagagatccctgggagaaatgcccgccgacatcttgggtcctcagatgagcctgcc 498
DB 294 gcacagagatccctgggagaaatgcccgccgacatcttgggtcctcagatgagcctgcc 353
QY 499 ctgtcctcgtccctgtgaggaaggaagacattagaaaatgaattgatgttctcttaa 558
DB 354 ctgtcctcgtccctgtgaggaaggaagacattagaaaatgaattgatgttctcttaa 413
QY 559 agdatgggcaggaacacagatcct 582
DB 414 agdatgggcaggaacacagatcct 437

RESULT 12
AAA06520/c
ID AAA06520 standard; cDNA; 301 BP.
XX
AC AAA06520;
XX
XX
DT 13-JUN-2000 (first entry)
XX
DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:287.
XX
KW *Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW immunogenic; cytostatic; vaccine; ss.
XX
OS Homo sapiens.
XX
XX
PN WO200004149-A2.
XX
XX
PD 27-JAN-2000.
XX
XX
PF 14-JUL-1999; 99WO-US15838.
XX
XX
PR 14-JUL-1998; 98US-0115453.
PR 14-JUL-1998; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX
XX
PA (CORI-) CORIXA CORP.
XX
XX
PI Dillon DC, Harlocker SL, Yugu J, Xu J, Mitcham JL;
XX
XX
DR WPI; 2000-171268/15.
XX
XX
PT New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumor protein -
XX
XX
PS Claim 1; Page 192; 263pp; English.
XX
XX
CC The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be

used in vaccines and in gene therapy. The polypeptides and
polynucleotides encoding them, antigen presenting cells which express
the polypeptides, antibodies against the polypeptides and vaccines
comprising them can be used for inhibiting the development of prostate
cancer in a patient. The polypeptides can be used to generate antibodies
or anti-idiotypic antibodies for passive immuno therapy. A portion of
the polynucleotides encoding the polypeptides can be used as a probe or
to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CAA82000 to AAY82020 represent sequences used in the exemplification of
the present invention.

XX
SQ Sequence 301 BP; 76 A; 58 C; 70 G; 97 T; 0 other;

Query Match 14.2%; Score 288.4; DB 21; Length 301;
Best Local Similarity 99.3%; Pred. No. 4,1e-73;
Matches 300; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 637 agcattaccaatgagaggaaaacagacagaaaattcttgatggcttcacaagacatgcaa 696
DB 301 AGCATTACCAATGAGAGGAAAACAGACGAGAAAATCTTGATGGCTTCAACAGACATGCAA 242
QY 697 caaacaaaatggaatactgtgatgacatgagcagcagcaagctggggagagataaccacg 756
DB 241 CAAACAAAATGGAATACTGTGATAACATGAGGCAGCAAGCTGGGAGGAGATAACCCAG 182
QY 757 ggcagagggctcaggaattctgacctgctgctaaactgctgctacataaccacaaatcatt 816
DB 181 GGGCAGAGGCTCAGGATTTGCCCCCTGCTGCTTAACCTGCTGCTTCAACCAAAATCAIT 122
QY 817 tcattatttcacccctcaaaacaaagctgtgttaatatctgatctctacggttcctctg 876
DB 121 TCATATTCTTAACCCCTCAAAACAAAGCTGTGTGAATATCTGATCTCTAC-GTTCTCTCG 63
QY 877 ggcacaaatcttcacatataccagccacactcatttttaatttagttccagatctg 936
DB 62 GGCCCAACATTCCTATATATCCAGCCACACTCATTTTAAATATTAGTTCCAGATCTG 3
QY 937 ta 938
DB 2 TA 1

RESULT 13
AAZ33445
ID AAZ33445 standard; cDNA; 359 BP.
XX
AC AAZ33445;
XX
DT 08-DEC-1999 (first entry)
XX
DE Human prostate cancer-associated EST 23.
XX
XX
KW Expressed sequence tag; EST; prostate tumor; antitumor; treatment;
KW gene therapy; tissue specificity human; ss.
OS Homo sapiens.
XX
XX
PN DE19811193-A1.
XX
XX
PD 16-SEP-1999.
XX
XX
PF 10-MAR-1998; 98DE-1011193.
XX
XX
PR 10-MAR-1998; 98DE-1011193.
XX
XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
XX
PI Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dahl E, Rosenthal A;
XX
XX
DR WPI; 1999-519628/44.
XX
XX
DR P-PSDB; AAY48243.

XX	AAF58254 standard; DNA; 936 BP.
AC	AAF58254;
DT	24-APR-2001 (first entry)
DE	Oligonucleotide D1875.
XX	Electron-transfer group; ETM; mismatch; genotyping;
KW	gene expression; ss.
QS	Synthetic.
PN	WO200107665-A2.
XX	01-FEB-2001.
XX	26-JUL-2000; 2000WO-US20476.
PR	26-JUL-1999; 99US-0145695.
PR	17-MAR-2000; 2000US-0190259.
XX	(CLIN-) CLINICAL MICRO SENSORS INC.
PA	Umek RM;
PI	UPI; 2001-159728/16.
DR	Nucleic acids containing electron-transfer group, useful as labels in hybridization assays, e.g. for genotyping, allowing repeat analyses on a single surface
XX	Example 6; Page 127; 159pp; English.
CC	The present invention relates to a composition comprising two nucleic acids each containing an electron-transfer group (ETM) having different redox potentials. The invention is used for electronic detection of nucleic acids, especially of substitutions (mismatches) and single-nucleotide polymorphisms, e.g. for genotyping, monitoring gene expression.
SS	Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;
SQ	
Query Match 6.7%; Score 136; DB 22; Length 936;	
Best Local Similarity 0.9%; Pred NO. 5.4e-29;	
Matches 7; Conservative 491; Mismatches 279; Indels 0; Gaps	
Qy	1132 aattctatcaacatcatcctcagtgcttggcccatcactgaaatttcattcccactt 1191 ::: : :: :
Db	777 WWWWWWMM 718 ::: : :: :
Qy	1192 ttgtgccattctcaagacctcaaatgtcattccatttaatacacaggattaccatttt 1251 :
.Db	717 WWWWWWMM 658 :
Qy	1252 tttttaacctggaagaattcaaagtTacaTgcagctatgggaatttaatatcatatttg 1311 :
Db	657 WWWWWWMM 598 :
Qy	1312 tttccagtgc aaagatgactaagtcctttatccctcccccttttgttgattttttcca 1371 : : : :
Db	597 WWWWWCMM 538 :
Qy	1372 gtataaggttaaaatgcttagcctgtactgaggcgctgtacacagcacagcctctccocat 1431 :
Db	537 WWWWWWMM 478 :
Qy	1432 ccctccagccttatctgtcatcaccatcaaacccctcccatnysaccttaaacaaaacttaa 1491

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 01:48:13 ; Search time 10902.7 Seconds
(without alignments)
2889.911 Million cell updates/sec

Title: US-09-402-713A-1
Perfect score: 2037
Sequence: 1 agaagctggcatcagaaaa.....caataaagaattacaaga 2037

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*

1: gb_bal:*

2: gb_ba2:*

3: gb_ba3:*

4: gb_in1:*

5: gb_in2:*

6: gb_in3:*

7: gb_om:*

8: gb_ov:*

9: gb_pat1:*

10: gb_pat2:*

11: gb_ph:*

12: gb_pl1:*

13: gb_pl2:*

14: gb_pl3:*

15: gb_pl4:*

16: em_ba1:*

17: em_ba2:*

18: em_fun:*

19: em_htgo_hum:*

20: em_htgo_inv:*

21: em_htgo_rod:*

22: em_htg_hum1:*

23: em_htg_hum2:*

24: em_htg_hum3:*

25: em_htg_hum4:*

26: em_htg_hum5:*

27: em_htg_hum6:*

28: em_htg_hum7:*

29: em_htg_hum8:*

30: em_htg_inv1:*

31: em_htg_inv2:*

32: em_htg_other:*

33: em_htg_rod:*

34: em_hum1:*

35: em_hum2:*

36: em_hum3:*

37: em_hum4:*

38: em_hum5:*

39: em_hum6:*

40: em_hum7:*

41: em_in:*

42: em_om:*

43: em_or:*

44: em_ov:*

45: em_pat:*

46: em_ph:*

47: em_pl:*

48: em_ro:*

49: em_sts:*

50: em_sy:*

51: em_un:*

52: em_vi:*

53: gb_sts1:*

54: gb_sts2:*

55: gb_sts3:*

56: gb_sy:*

57: gb_un:*

58: gb_vil:*

59: gb_vl2:*

60: gb_htg1:*

61: gb_htg2:*

62: gb_htg3:*

63: gb_htg4:*

64: gb_htg5:*

65: gb_htg6:*

66: gb_htg7:*

67: gb_htg8:*

68: gb_htg9:*

69: gb_htg10:*

70: gb_htg11:*

71: gb_htg12:*

72: gb_htg13:*

73: gb_htg14:*

74: gb_htg15:*

75: gb_htg16:*

76: gb_htg17:*

77: gb_htg18:*

78: gb_htg19:*

79: gb_htg20:*

80: gb_htg21:*

81: gb_htg22:*

82: gb_htg23:*

83: gb_htg24:*

84: gb_htg25:*

85: gb_pr1:*

86: gb_pr2:*

87: gb_pr3:*

88: gb_pr4:*

89: gb_pr5:*

90: gb_pr6:*

91: gb_pr7:*

92: gb_pr8:*

93: gb_pr9:*

94: gb_ro1:*

95: gb_ro2:*

96: gb_in4:*

97: gb_pr10:*

98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	2032	99.8	3923	88	AF103907 Homo sapi
2	1589.4	78.0	5435	88	AF103908 Homo sapi
3	1555.8	76.4	164371	80	AL390239 Homo sapi
4	1361.4	66.8	173831	80	AL359314 Homo sapi
5	1317.8	64.7	267581	80	AL358573 Homo sapi
6	235.4	11.6	359	9	AX018075 Sequence
7	99	4.9	143675	79	AL161625 Homo sapi
8	99	4.9	267581	80	AL358573 Homo sapi

```
9      98      4.8 580 89 AF279290 Homo sapi
c 10 97.4 4.8 172298 61 AC009556
11 59.8 2.9 7218 10 I66494
c 12 57.4 2.8 7218 10 I66494
13 52.6 2.6 1141 10 AX083744
14 47.2 2.3 137064 74 AC073363
c 15 47.2 2.3 152571 66 AC021589
c 16 47.2 2.3 160133 69 AC025140
c 17 47.2 2.3 167574 62 AC011274
c 18 47.2 2.3 177396 74 AC069219
c 19 47.2 2.3 189915 70 AC028861
c 20 47.2 2.3 191034 80 AL356372
c 21 47 2.3 10968 9 AR029638
c 22 47 2.3 10968 14 LEU65391
c 23 47 2.3 59202 13 AF220603
c 24 47 2.3 60831 13 AF220602
c 25 45.8 2.2 188781 67 AC022546
c 26 45.8 2.2 323479 77 AC087779
c 27 45.6 2.2 40873 83 CEH04109
c 28 45.6 2.2 42141 6 CEY54G9A
c 29 45.6 2.2 300197 83 CEY54G9
c 30 45.4 2.2 5134 9 AR085492
c 31 45.4 2.2 5475 9 AR029637
c 32 45.2 2.2 1141 10 AX083744
c 33 45 2.2 191319 63 AC015653
c 34 44.4 2.2 151553 90 AL139342
c 35 44.4 2.2 181545 69 AC025529
c 36 44.2 2.2 163958 4 AC010705
c 37 44 2.2 106935 6 CELY37E3
c 38 43.8 2.2 171471 64 AC015844
c 39 43.2 2.1 120984 86 AC006463
c 40 43.2 2.1 184316 68 AC024022
c 41 42.8 2.1 3269 2 AF270032
c 42 42.8 2.1 9131 85 AB038781
c 43 42.6 2.1 930 53 CNS06TZZ
c 44 42.6 2.1 8912 85 AB038782
c 45 42.6 2.1 10708 89 AF113616
```

ALIGNMENTS

```
RESULT 1
AF103907 AF103907 3923 bp mRNA PRI 14-AUG-2000
LOCUS Homo sapiens non-coding RNA DD3 sequence.
ACCESSION AF103907
VERSION AF103907.1 GI:6165973
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3923)
AUTHORS Bussemakers,M.J., van Bokhoven,A., Verhaegh,G.W., Smit,F.P.,
Karthaas,H.F., Schalken,J.A., Debruyne,F.M., Ru,N. and Isaacs,W.B.
DD3: a new prostate-specific gene, highly overexpressed in prostate
cancer
JOURNAL Cancer Res. 59 (23), 5975-5979 (1999)
MEDLINE 20072260
PUBMED 10606244
REFERENCE 2 (bases 1 to 3923)
AUTHORS Bussemakers,M.J.G., Van Bokhoven,A., Verhaegh,G.W., Smit,F.P.,
Karthaas,H.F.M., Schalken,J.A., Debruyne,F.M.J., Ru,N. and
Isaacs,W.B.
Direct Submission
JOURNAL Submitted (28-OCT-1998) Urology Research Laboratory, University
Hospital Nijmegen, P.O. Box 9101, Nijmegen 6500 HB, Netherlands
FEATURES
Location/Qualifiers
1..3923
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
```


Db 863 AGCTGTTGTAATATCTGATCTCTACGGTTCCTCTGGGCCCAACATTCCTCATATATCCA 922
Qy 901 gccacactcattttaataatttagtccagatctgtactgtgaccttcttacactgtag 960
Db 923 GCCACACTCATTTTAAATATTTAGTTCCAGATCTGTACTGTGACCTTTCTACACTGTAG 982
Qy 961 aatacattactcatttggttcaagacccttctgtgtgcctcaataatgtagctgact 1020
Db 983 AATAACATTACTCATTTGTTCAAGACCCCTCGTGTGCGCTAAATATGTAGCTGACT 1042
Qy 1021 gtttttccaaaggagtggtcttgcccgagggatctgtgaacaggctgtggaagcatctcaa 1080
Db 1043 GTTTTTCCTCAAGGAGTGTCTGGCCAGCGGATCTGTCAACAGGCTGGGAAGCATCTCAA 1102
Qy 1081 gatcttccagggttatacttactacagacacagcatgatacaggagtgaattatcta 1140
Db 1103 GATCTTTCCAGGGTTATCTTACTAGCACAGCATGATCATPACGGAGTGAATTTATCTA 1162
Qy 1141 atcaacatcctcctcagtgctcttgcccatactgaaattcattccocacttttggccca 1200
Db 1163 ATCAACATCATCCTCAGTGTCTTTGCCCATACTGAAATTCATTTCCCACTTTTGTGCCCA 1222
Qy 1201 ttctcaagacctcaaaatgtcattccatttaataatcacaggattaaacttttttttaacc 1260
Db 1223 TTCTCAAGACCTCAAAATGTCTATCCATTAATATCACAGGATTAATCTTTTAAACC 1282
Qy 1261 tggagaataatcaatgttccatgcagctgagtgatggaatttaatacatatttgtttccagt 1320
Db 1283 TGGAGAATCAATGTGATGACATGACATGAGGAAATTAATACATATTTGTGTTCCAGT 1342
Qy 1321 gcaagatgactaaagcttctatccctccctctgttggatttttttccagtataaagt 1380
Db 1343 GCAAAGATGACTAAGTCTTTATCCCTCCCTTGTGTTGATTTTTTCCAGTATAAAGT 1402
Qy 1381 taaaatgcttgactgtactgagtgctgtatcacgacagcctctcccatccctccagc 1440
Db 1403 TAAATGCTTAGCCCTGTACTGAGGCTGTATACACACAGCGCTCTCCCACTCCCTCCAGC 1462
Qy 1441 ctatctctcatcaacatcaacccctccatnysacctaaacaaatactaacttgaatt 1500
Db 1463 CTATCTGTCTATCACCATCAACCCCTCCATACACACCTTAAACAAATCTAATCTGTAAT 1522
Qy 1501 ccttgaacatgctcagncatacattttcctctcctgcctgagaagctctctctgtcctt 1560
Db 1523 CCTGAAACATGTCAGGACATACATTTCCCTCTGCTGAGAGCTCTTCCCTGTCTCTT 1582
Qy 1561 aantctagaatgatgaaagtttgaataagtgtgactatcttacttcatgcaaaagagg 1620
Db 1583 AAATCTAGAATGATGTAAGTGTGTAATGAGTTGACTATCTTACTTCATGCAAGAAGGG 1642
Qy 1621 acacatatgagattcatcatcatcagacagcaaaatacctaaagtgaatttgattata 1680
Db 1643 ACACATATGAGATTTCATCATCATGAGACAGCAAAATCTAAAGTGAATTTGATTATA 1702
Qy 1681 agagtttagataaataatgaaatgaagacacagaggaatgtttatggggcacgtt 1740
Db 1703 AGAGTTTGAATAAATATATGAATGCAGAGCCACAGAGGGAATGTTTATGGGACCGTT 1762
Qy 1741 tgaagcctggagtgatgaaagagcagggaacctcatagttatcttataataatactt 1800
Db 1763 TGAAGCCTGGGATGTGAAGCAAGGAGGGAACCTCATAGTATCTTATATAATATACTT 1822
Qy 1801 catttctctatctctatcacaaataccaaagcttttcaagaattcatgcagtgcaaa 1860
Db 1823 CATTTCTCTATCTCTATCACATATCCAAACAGCTTTTTCACAGAATTCATGCAGTGCAAA 1882
Qy 1861 tcccaaaaggtaaacctttatccatttcatggtgagtcgccttagaattttggcaaatca 1920
Db 1883 TCCCAAAAGGTAACCTTTATCCATTTTCATGTTGAGTGGCTTTAGAAATTTGGCAATCA 1942
Qy 1921 tactggtcacttatctcaacttggatgtgttgcctttagtttaattgaaagaaata 1980

Db 1943 TACTGGTCACCTTATCTCAACTTTGAGATGTGTTTCTCCTGTAGTTAATTTGAAAGAAATA 2002
Qy 1981 gggcactctttagcaccactttaggttccactctctgcaataaagaatttacaaga 2037
Db 2003 GGGCACCCTTTGTGAGCCACTTTAGGGTTCACTCTCTGGCAATAAAGAATTTACAAGA 2059
RESULT 2
AF103908 Locus AF103908 5435 bp DNA PRI 14-AUG-2000
DEFINITION Homo sapiens non-coding RNA DD3 gene, exons 2, 3, and 4.
ACCESSION AF103908
VERSION AF103908.1 GI:6165974
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 5435)
AUTHORS Bussemakers,M.J., van Bokhoven,A., Verhaegh,G.W., Smit,F.P.,
Karthaas,H.F., Schalken,J.A., Debruyne,F.M., Ru,N. and Isaacs,W.B.
DD3: a new prostate-specific gene, highly overexpressed in prostate
cancer
JOURNAL Cancer Res. 59 (23), 5975-5979 (1999)
MEDLINE 20072260
PUBMED 10606244
REFERENCE 2 (bases 1 to 5435)
AUTHORS Bussemakers,M.J.G., Van Bokhoven,A., Verhaegh,G.W., Smit,F.P.,
Karthaas,H.F.M., Schalken,J.A., Debruyne,F.M.J., Ru,N. and
Isaacs,W.B.
TITLE Direct Submission
JOURNAL Submitted (28-OCT-1998) Urology Research Laboratory, University
Hospital Nijmegen, P.O. Box 9101, Nijmegen 6500 HB, Netherlands
FEATURES
source 1..5435
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/map="9q21-q22"
repeat_region 1..78
/rpt_family="Alu"
/rpt_type="dispersed"
exon 533..697
/note="Alternative exon present in 5% of cDNA clones"
repeat_region 1035..1294
/rpt_family="Alu"
/rpt_type="dispersed"
mRNA join(1571..1753,1981..5435)
/note="transcript III"
mRNA /product="non-coding RNA DD3"
join(1571..1753,1981..2517)
/note="transcript I"
mRNA /product="non-coding RNA DD3"
join(1571..1753,1981..3579)
/note="transcript (major) II"
exon /product="non-coding RNA DD3"
1571..1753
/number=3
exon 1981..5435
/number=4
polyA_signal 2495..2499
/note="transcript I"
polyA_site 2517
/note="transcript I"
polyA_signal 3553..3558
/note="transcript (major) II"
polyA_site 3579
/note="transcript (major) II"
repeat_region 5268..5423
/note="LINE"
/rpt_family="L1"
/rpt_type="dispersed"

polyA_site 5435
BASE COUNT 1546 a 1188 c 1036 g 1654 t 11 others
ORIGIN

Query Match 78.0%; Score 1589.4; DB 88; Length 5435;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1587; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 442 cagagatccctgggagaaatgcccggcccccattcttggttcacatgcagctgcgcctg 501
DB 1976 CACAGATCCTGGGAGAAATGCCGCCGCCATCTTGGGTCTCGATGAGCCTCGCCCTG 2035

QY 502 tgcctgtccctgtgaggaagacattagaaaatgaattgatgtgttcttaag 561
DB 2036 TGCCTGTCCCGCTGTGGAGGAGACATAGAAAATGAATGTGTTCTCTTAAGG 2095

QY 562 atgggcagaaacagatcctgtgtggatatttttgaacgggattcacagatttga 621
DB 2096 ATGGGCAGGAAACAGATCCTGTGTGGATATTTATTTGAACGGGATTTACAGATTTGAAA 2155

QY 622 tgaagtcacaagtgcattaccaatgagagaaacagagagaaaatcttgatgct 681
DB 2156 TGAAGTCACAAAAGTGAGCATTTACCAATGAGAGGAAACACAGAGAAAATCTTGATGGCT 2215

QY 682 tcacaagacatgcacaacaaatggaatactgtgatgacatgagcagcagcaagctgg 741
DB 2216 TCACAGACATGCAACAACAAATGGAATGATGTGATGACATGAGCGACGCAAGCTGGG 2275

QY 742 gaggagataaacacggggcaggggtcaggattctggccctgcctcaaacctgtgcgt 801
DB 2276 GAGGAGATAACACGGGCGAGGGGTGAGGATTTCTGGCCCTGCTGCTTAACTGTGCGTT 2335

QY 802 cataaaccaatattcatatttctaaccctcaaaacaaagctgtgttaatactgatct 861
DB 2336 CATAAACCAATCATTTCAATTTCTTAACCTCAAAAACAAAGCTGTGTGAATATCTGATCT 2395

QY 862 ctacggttctcttgcccaacattctccatatatccagccacacactatttttaatt 921
DB 2396 CTACGGTCTCTTGCGCCCAACATTTCCATATATCCAGCCACACACTCATTTTATATT 2455

QY 922 tagttccacagatctgtactgtgacctttcacactgtagaataaacattactcattttgt 981
DB 2456 TAGTTCACAGATCTGTACTGTGACCTTTCTACACTGTAGAATAAACATTACTCATTTGTT 2515

QY 982 caaagaccctctgtctgcctaataatgtagctgaactgttttcttaaggatgtct 1041
DB 2516 CAAGAGCCCTTCGTGTGCTGCTTAATATGTAGCTGACTGTGTTTCTTAAGGAGTGTCT 2575

QY 1042 ggcacagggatctgtgaacagctgggaagcatctcaagatctttccagggtataact 1101
DB 2576 GGCACAGGGGATCTGGAACAGCTGGGAAGCATCTCAAGATCTTTCCAGGGTTATACTT 2635

QY 1102 actagcacacagatgatcattacggagtgaaattatctaatcaaatcatcctcagtgte 1161
DB 2636 ACTAGCACACAGCATGATCATTAACGGAGTGAATATTAATAACATCATCTCCTCAGTGTC 2695

QY 1162 ttggccatactgaaattcttccacttttgtgcccattctcaagacctcaaaatgtc 1221
DB 2696 TTGGCCATCTACGAAATTCATTTCACATTTTGTGCCCCATTCTCAAGACCTCAAAATGTC 2755

QY 1222 attccattaatcacaggatcaacttttttttaacctggaagaattcaagtgttaact 1281
DB 2756 ATTCCATTATATCACAGGATTAACCTTTTAACTTGAACCTGGAAGAAATTCATGTTACAT 2815

QY 1282 gaagctatgggaatttaatacatattttgtttccagtcgaagaatgactgaagtccttt 1341
DB 2816 GCAGCTATGGGAATTTAATATATTTTGTGTTTCCAGTGCAAGATGACTAAGTCCTTT 2875

QY 1342 atccccctctgtttgttttttttccagtataaagttaaaatgcttagcctgtact 1401
DB 2876 ATCCCCCTCTGTGTTGATTTTTCAGGATAAAGTTAAATGCTTAGCCTGTGACT 2935

QY 1402 gaggctgtatcacagcacagcctctcccccattccctccagccttatctgtcatcaccatcaa 1461
DB 2936 GAGGCTGTATACAGCACACAGCCTCTCCCATCCTCCAGCCTTATCTGTCTATCACCATCAA 2995

QY 1462 cccctcccatnysacctaaacaaatcttaactgttaattccttgaacatgtcaggncata 1521
DB 2996 CCCCTCCCATACCAACCAATAAATACTAATCTGTAATTTCTTGAACATGTGAGGACATA 3055

QY 1522 caatttcccttgcctgagaagccttctctgtctcttaantctagaatgtataaagt 1581
DB 3056 CATTAATCTCTGCTTGAGAAAGCTTCTCTTGTCTCTTAAATCTAGAATGATGTAAGAT 3115

QY 1582 ttgtaataagttgacttatcttctcatcgaagagggacacatatagattcattcatc 1641
DB 3116 TTTGAATAAGTTGACTATCTTACTTCATCAAGAAGGACACATATGAGATTTCATCATC 3175

QY 1642 acatgagacagcaaatactaaagttaatttgattataagagtttagataaataataga 1701
DB 3176 ACATGAGACAGCAAAATACTAAAAGTGAATTTGATTATAAGAGTTTAGATAAATATATGA 3235

QY 1702 aatgcaagkccacagaggggaatgtttatgggcacgtttgtaagcctgggatggaagm 1761
DB 3236 AATGCAAGAGCCACAGAGGGAATGTTTATGGGCGACGTTTGTAAAGCTGGGATGTGAAGC 3295

QY 1762 aaaggcaggaacccctcatcttatataataataacttcttctctctctctatcaca 1821
DB 3296 AAAGCGAGGAACCTCATAGTATCTTATATATATATATCTTCTATCTCTATCTATCACA 3355

QY 1822 atatacaacaaagcttttccagaattcatgcagtgcaaatccccaaagtaaccccttate 1881
DB 3356 ATATCAACAAGCTTTTTCAGAAATTCATGTCAGTGCAGTGCAGAAATCCCAAGGTAACCTTTATC 3415

QY 1882 catttcagtgtagtgcgcttttagaattttggcaaatcatctactgttcaacttatctcaact 1941
DB 3416 CATTTTCATGGTGAGTGGCGCTTTAGAAATTTTGGCAATCATACTGTGTCACCTTTATCTCAACT 3475

QY 1942 ttgagatgtgttctctctgttagttaattgaaagaaataggcactcttctgtgagccact 2001
DB 3476 TTGAGATGTGTTGCTTGTGTTGTTAGTTAATTTGAAGAATAAGGCGACCTCTTGTGAGCCACTT 3535

QY 2002 tagggttcaactcctggcaataaagaatttacaaga 2037
DB 3536 TAGGGTTCACTCTCTGGCAATAAAGAATTTACAAAGA 3571

RESULT 3
AL390239 164371 bp DNA HTG 09-MAR-2001
LOCUS Homo sapiens chromosome 9 clone RP11-58J3, *** SEQUENCING IN
DEFINITION PROGRESS ***, 4 unordered pieces.
ACCESSION AL390239
VERSION AL390239.11 GI:13274794
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 164371)
Burton, J.
Direct Submission
JOURNAL Submitted (05-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Mar 12, 2001 this sequence version replaced gi:13273805.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA58J3
----- Summary Statistics

Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 163337 bases at least Q40
 Consensus quality: 163481 bases at least Q30
 Consensus quality: 163577 bases at least Q20
 Insert size: 164071; sum-of-contigs
 Insert size: 166918; 4.2% error; agarose-1p
 Quality coverage: 8.78x in Q20 bases; sum-of-contigs Quality
 coverage: 9.07x in Q20 bases; agarose-1p

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 20240: contig of 20240 bp in length
 * 20241 20340: gap of 100 bp
 * 20341 116100: contig of 95760 bp in length
 * 116101 116200: gap of 100 bp
 * 116201 118243: contig of 2043 bp in length
 * 118244 118343: gap of 100 bp
 * 118344 164371: contig of 46028 bp in length.

FEATURES

Source

1. 164371
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="9"
 /clone="RP11-58J3"
 /clone.lib="RPC1-11.1"
 1. 20240
 /note="assembly_fragment:00838
 fragment_chain:1
 clone_end:SP6
 vector_side:left"
 20341..116100
 /note="assembly_fragment:01998
 fragment_chain:1"
 116201..118243
 /note="assembly_fragment:00916
 fragment_chain:1"
 118344..164371
 /note="assembly_fragment:00334
 fragment_chain:1
 clone_end:T7
 vector_side:right"
 50254 a 34905 c 32663 g 46249 t 300 others

BASE COUNT
ORIGIN

Query Match 76.4%; Score 1555.8; DB 80; Length 164371;
 Best Local Similarity 99.4%; Pred. No. 0;
 Matches 1587; Conservative 4; Mismatches 3; Indels 3; Gaps 3;
 QY 442 cagagatccctggagaaatgccgcgcacatcttgggtcagatgagcctgccttg 501
 DB 33316 CACAGATCCCTGGGAGAAATCCCGGCCCATCTTGGGTCTATCGATGAGCCTCGCCCTG 33375
 QY 502 tgccctggctccctgtgaggaagagacattagaaataaattgatgtgtctcttaaagg 561
 DB 33376 TGCCCTGGTCCCTGTGAGGAAGGACATTAGAAAATGAATTGATGTGCTCTTAAGG 33435
 QY 562 atgggcaggaacacagatcctgttggatattttggaacgggattacagatttgaa 621
 DB 33436 ATGGGCAGGAAAACAGATCCTCTGTGGGATATTTATTTGAACGGGATTACAGATTGAAA 33495
 QY 622 tgaagtcacaaagttagcattaccatagagagaaacacagacagaaaattcttgatgct 681
 DB 33496 TGAAGTCAAAAAGTGAGCATTCACCAATGAGAGAAACAGACGAGAAAATCTTGATGGCT 33555

QY 682 tcacaagacatgcaacaaacaaatggaaatctgtgatgacatgaggcagccaaactggg 741
 DB 33556 TCACAAGACATGTCACAAACAAATGGAATACTGTGATGACATGAGGCAGCAAGCTGGG 33615
 QY 742 gaggagataacacacagggcagaggctcaggattcttgccctgctgcctaaactgtgcgtt 801
 DB 33616 GAGGAGATAACCAACGGGCAGAGGCTCAGGATCTTGCCCTGCTGCTTAAACTGTGCGTT 33675
 QY 802 cataacaaaatcattctcatttctaaacctcaaaacaaagctgtgttaatatctgatct 861
 DB 33676 CATACACAAATCATTTTCATATTTCTAACCTCAAAACAAAGCTGTGTAATATCTGATCT 33735
 QY 862 ctacggttcctctgggcccacaattctccatatatccacgcccacacactcatctttaaatt 921
 DB 33736 CTACGGTTCTTCTGGGCCCAACATTCCTCATATATCCAGGCACACACTCATTTTAAATTT 33795
 QY 922 tagttccacagatctgtactgtgacaccttctacactgtagaataacataactactatttctt 981
 DB 33796 TAGTTCCACAGATCTGTACTGTGACCTTTCTACACTGTAGATAAACATTACTCATTTTGT 33855
 QY 982 caaagacacctctgtgtctgctcctaataatgtagctgactgttttctcctaagagtgctct 1041
 DB 33856 CAAAGACCTTCTGCTGCTTAATATGATGACTGACTGTTTTTCTTAAGGAGTGTTCT 33915
 QY 1042 ggcacagggatctgtgaacagggctgggaagcatctcaagatccttccaggggttaactct 1101
 DB 33916 GSCCCAGGGGATCTGTGAACAGGCTGGGAAGCATCTCAAGATCTTTCCAGGGTTTACTTT 33975
 QY 1102 actagcacacagcatgatacttaacggagtgaaattatctataacacatcatctcctcagtgct 1161
 DB 33976 ACTAGCACACAGCATGATCATTTACGGAGTGAATATCTAATAACATCATCTCCTCAGTCTC 34035
 QY 1162 ttgcccatactgaaatcattctccacctttgtgcccattctcgaagacctcaaaatgctc 1221
 DB 34036 TTGCCCATACTGAAATCAITTCCTCCACTTTTGCCCCATTCTCAAGACCTCAAAATGTC 34095
 QY 1222 attccattaatatcacagattaaacttttttttaacctggaaataatcaattgttatcat 1281
 DB 34096 ATTCCATTAAATATCACAGGATTAACTTTTATTTTAACTGGAAGATTTCAATGTTTACAT 34155
 QY 1282 gcagctatgggaatttaattacataattttgtttccagtgcaaaagatgaactaagctccttt 1341
 DB 34156 GCAGCTATGGGAATTAATATACATATTTTGTTCAGTGCAGAAAGTACTAAGTCCCTTT 34215
 QY 1342 atccctccctttgttgaatttttttccagataaagctaaatgctcctgactgact 1401
 DB 34216 ATCCCTCCCTTTGTTGATTTTTTTTCCAGATATAAAGTTAAATGCTTAGCCTTGACT 34275
 QY 1402 gaggctgtatcacag-cacagcctctcccacatccctcccagccttatctgtcatcacatca 1460
 DB 34276 GAGGCTGTATACAGCCACAGCCTCTCCCCATCCCTCCAGCCTTATCTGTCTCATCACATCA 34335
 QY 1461 accctcccatacysacctaaacaaatctaatctgtaattccttgaacatgtcaggnact 1520
 DB 34336 ACCCTCCCATG-CACCTAAACAAATCTAATCTGTAATTCCTTGAACATGTCAGG-CAT 34393
 QY 1521 acatrttctctcctgcctgcagagcctcttccctgtctcttaantctagaatgatgaaag 1580
 DB 34394 ACATATTCTCTTCCCTGCCTGAGAGCTCTTCTCTCTTAANTCTAGAATGATGTAAG 34453
 QY 1581 ttttgaataagtgtactacttcttactctcatgcaaaagagggacacatatagagattcatcat 1640
 DB 34454 TTTTGAATAAGTTGACTATCTTACTTCATGCAAGAGGACACATATGAGATTTCATCAT 34513
 QY 1641 cacatgagacagcaataactaaaagttaatttgattataagagtttagataaaatatatg 1700
 DB 34514 CACATGAGACAGCAATACTAAAAAGTGAATTTGATTTATAAGAGTTTAGATAAATATATG 34573
 QY 1701 aaatgcagagkccacagaggggaattttatggggcacgttttgaagcctgggatggaag 1760
 DB 34574 AAATGCAAGAGCCACAGAGGGGAATGTTTATGGGGCACGCTTTGTAAGCCCTGGGATGGAAG 34633
 QY 1761 maaaggcagggaacctcatagtatcttataataataacttctctctctctctctctc 1820

```

Db 34634 CAAAGCAGGGAACCATAGTATCTTATATAATATACTTCATTTCTATCTCTATCAC 34693
QY 1821 aatatccaaagctttcacagaattcatgagtcagtaacatcccaagtaacctttat 1880
Db 34694 AATATCCAAACAGCTTTTACAGAAATTCATGAGTCAGTCAATCCCAAGGTAACTTTAT 34753
QY 1881 ccatttcagtgtgagtcgcttttagaattttggcaaatcatactatgctcattatctcaac 1940
Db 34754 CCATTTCATGGTGAGTCGCTTTAGAAATTTGGCAATCATATGCTGCTACTTATCTCAAC 34813
QY 1941 tttagatgtgttgcctttagttaaattgaagaataaggacactctgtgagccact 2000
Db 34814 TTTGAGATGTGTTGTCTGTAGTTAATGAAAGAAATAGGCACTCTGTGAGCCACT 34873
QY 2001 ttagggttcactctgagcaataaagaattacaaga 2037
Db 34874 TTAGGGTTTCACTCCTGGCAATAAGAAATTTACAAGA 34910

```

```

RESULT 4
AL359314 173831 bp DNA HTG 08-APR-2001
LOCUS Homo sapiens chromosome 9 clone RP11-108L4, *** SEQUENCING IN
DEFINITION PROGRESS ***, 3 unordered pieces.
ACCESSION AL359314
VERSION AL359314.12 GI:13396560
KEYWORDS HTG: HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 173831)
Leongamornlert, D.
Direct Submission
Submitted (07-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requests: clonerequest@sanger.ac.uk
On Mar 20, 2001 this sequence version replaced gi:13277120.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: ba108L4
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 173545 bases at least Q40
Consensus quality: 173599 bases at least Q30
Consensus quality: 173621 bases at least Q20
Insert size: 173631; sum-of-contigs
Insert size: 172123; 10.0% error; agarose-fp
Quality coverage: 10.35x in Q20 bases; sum-of-contigs Quality
coverage: 10.50x in Q20 bases; agarose-fp
-----

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 50595: contig of 50595 bp in length
* 50596 50695: gap of 100 bp
* 50696 157636: contig of 106941 bp in length
* 157637 157736: gap of 100 bp
* 157737 173831: contig of 16095 bp in length.
* Location/Qualifiers
1..173831

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

* 1 50595: contig of 50595 bp in length
* 50596 50695: gap of 100 bp
* 50696 157636: contig of 106941 bp in length
* 157637 157736: gap of 100 bp
* 157737 173831: contig of 16095 bp in length.
* Location/Qualifiers
1..173831

```

```

FEATURES
source
1..173831

```

```

/misc_feature
1..50595
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-108L4"
/clone_lib="RPC1-11.1"
/feature="assembly_fragment:04137"
fragment_chain:1"
50696..157636
/feature="assembly_fragment:04236"
fragment_chain:1"
157737..173831
/misc_feature
/feature="assembly_fragment:01656"
BASE COUNT 52107 a 36469 c 35018 g 50037 t 200 others
ORIGIN

```

```

Query Match 66.8%; Score 1361.4; DB 80; Length 173831;
Best Local Similarity 92.4%; Pred. No. 0;
Matches 1395; Conservative 4; Mismatches 107; Indels 3; Gaps 3;
QY 442 cagagatccctgggagaaaatgccggccgcacatcttgggtcgcagatgagcctgcctgtg 501
Db 156237 CACAGATCCCTGGGAGAAATGCCGGCCGCCATCTTGGGTCTATCGATGAGCCTGCCCTG 156296
QY 502 tgcctgggtcccgcttggaggaagacattagaaaaatgaattgattgttctttaaagg 561
Db 156297 TGCTGGTCCCCTTGTGAGGAGAGACATTAGAAAATGAATGATGTCTTCTTAAAGG 156356
QY 562 atgggcaggaacacagatcctgttgggataattttgaacgggattacagatttgaa 621
Db 156357 ATGGGCAGGAAACAGATCCTGTGTGGGATATTTATTTGAACGGGATTACAGATTGAAA 156416
QY 622 tgaagtcacaaagtgcagcattaccgaatgagagaaaacagagagaaaatcttgagct 681
Db 156417 TGAAGTCACAAAGTGGAGCATTACCAATGAGAGGAAACAGAGAGAAAATCTTGATGGCT 156476
QY 682 tcacaagcatgcacaaacaaatggaatactgtgacacagagcagcagcaagctggg 741
Db 156477 TCACAAGACATGCACAAACAAATGGAATAGTGTGATGACATGAGGCAGCCAGCTGGG 156536
QY 742 gaggagataaccacgggagcaggggtcaggattctggccctgcctcctaaactgtgcgtt 801
Db 156537 GAGGAGATAACCCAGGGGAGAGGGTCAAGATTCTGGCCCTCTGCTTAACTTGTGCTT 156596
QY 802 catacacaatcatctcatatttcaacctcaaacaaagctgttgtaatactgact 861
Db 156597 CATACCAAAATCATTTATATTTCTAACCTCTAAACAAAGCTGTTGTAATATCTGATCT 156656
QY 862 ctacggttctcttggcccaacattctccatatatccagccacacactatcttataatt 921
Db 156657 CTACGGTTCTTCTGGGCCCAACATTCTCCATATATCCAGCCACACTCATTTTAAATAT 156716
QY 922 tagttccagatctgtactgtgacctttctacactgtagaataacacttactcttctgtt 981
Db 156717 TAGTTCCAGATCTGTACTGTGACCTTTCTACACTGTAGATAAATTAATTAATTTGTT 156776
QY 982 caaacacctctgtctgcctaataatgtagctgactgttttcttaagagctgtct 1041
Db 156777 CAAGACCCCTTCTGTCTGCTCTATATATGATGAGCTGACTGTTTCTTAAAGAGTGTCT 156836
QY 1042 gggccaggggactgtgaacaggctgggaagcatctcaagatcttccagggttatactt 1101
Db 156837 GGGCCAGGGGATCTGTGAACAGGCTGGGAAGCATCTCAAGATCTTCCAGGCTTATACCT 156896
QY 1102 actagcacacagcatgatactacagagtgatattatctaatcaacatactcctcagtgct 1161
Db 156897 ACTAGCACACAGCATGATCATTTACGGAGTGAATTTATCTAATCAACATCATCTCAGTGC 156956
QY 1162 ttgcccactatgaattcatttccacctttgtgccactttgtgccattctcaagacctcaaatgtc 1221
Db 156957 TTTGCCCATACATGAATTCATTTCCACCTTTTGTGCCCATCTTCAAGACCTCAAAATGTC 157016

```

```
QY 1222 attccattaataacacagaggtaaaccttttttttaaacctggaagaattcaattgttacat 1281
|||||
Db 157017 ATTCATTAATATACAGAGTAACCTTTTTTTTAACTGGAAGATTCATGTTACAT 157076
|||||
QY 1282 cgagctatgggaatttaatacatatttggttttccagtgcaaaagatgaactaacgtcttt 1341
|||||
Db 157077 GCAGTATGGGAATTAATACATATTTTGTGTTTCCAGTGCAGAAATGACTAAGTCCTTT 157136
|||||
QY 1342 atccctccctctgtttgattttttccagttataaagttaaaatgcttgactgtact 1401
|||||
Db 157137 ATCCCTCCCTCTGTTGTTGATTTTTTTTCCAGTATAAAGTTAAATGCTTAGGCTGTACT 157196
|||||
QY 1402 gagggtgtatacag-cacagcctctcccccattccctccagccttattgtcatcacacatca 1460
|||||
Db 157197 GAGGCTGTATACAGGCACAGCCTCTCCCATCCCTCCAGCCTTATCTGTCATCACATCA 157256
|||||
QY 1461 accctcccatnysacctaaacaaactaacttgtaattccttgtaacatgtcaggnatc 1520
|||||
Db 157257 ACCCTCCCATG-CACCTAAACAAATCTAATCTGTAATTCCTTGAACATGTCAGG-CAT 157314
|||||
QY 1521 acattttctctctgcctgagagcctctctctgtctcttaantctagaatgatgaag 1580
|||||
Db 157315 ACATTTATCTCTGCTGCGTGAGAGCTCTCTCTGTCTCTTAATCTAGAATGATGTAAG 157374
|||||
QY 1581 ttttgaataagttgactatcttacttctcatgcaaaagagacacatatgagattcatcat 1640
|||||
Db 157375 TTTTGAATAAGTTGACTATCTTACTTCTATGCAAGAGGACACATATGAGATTCTATCAT 157434
|||||
QY 1641 cacatgagacagcaataactaaaagtgaattgtattgattataagagtttagataatatg 1700
|||||
Db 157435 CACATGAGACAGCAATAACTAAAGTGTAATTTGATTAATTAAGAGTTTAGATAAATATATG 157494
|||||
QY 1701 aaatcaagacacacagaggaagtgtttatgggcagcttttgaagcctgggatgtgaag 1760
|||||
Db 157495 AAATGCAAGAGCCACAGAGGGGAATGTTTATGGGCGACGTTTGTAGCCCTGGGATGTGAAG 157554
|||||
QY 1761 maaagcagggaacctcatagtagtcttatataataactatctctctctctctctcatc 1820
|||||
Db 157555 CAAAGCAGGGAACCTCATAGTATCTTATATAATATCTTCTATCTCTATCTATCTAC 157614
|||||
QY 1821 aaatcccaacaagcttttccacagaattcatcgagtgcaaatccccaaaggaacctttat 1880
|||||
Db 157615 AATATCCAAACAGCTTTTCACANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 157674
|||||
QY 1881 ccatttcagtgtagtgcgctttagaatttggcaaatcatactggtcacttctctcaac 1940
|||||
Db 157675 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 157734
|||||
QY 1941 tttagatg 1949
|||||
Db 157735 NNACACAAG 157743

RESULT 5
LOCUS AL358573/c
DEFINITION Homo sapiens chromosome 9 clone RP11-133022, *** SEQUENCING IN
PROGRESS ***, 37 unordered pieces.
ACCESSION AL358573
VERSION AL358573.17 GI:13660951
KEYWORDS HTG; HTGS-PHASE1; HTGS-DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 267581)
Direct Submission
Mclay,K.
Submitted (14-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Apr 17, 2001 this sequence version replaced gi:133988774.
----- Genome Center
```

```
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA133022
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 256599 bases at least Q40
Consensus quality: 259744 bases at least Q30
Consensus quality: 261407 bases at least Q20
Insert size: 263981; sum-of-contigs
Quality coverage: 5.51x in Q20 bases; sum-of-contigs Quality
Coverage: 12.23x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 25718: contig of 25718 bp in length
*
* 25719 25818: gap of 100 bp
*
* 25819 29567: contig of 3749 bp in length
*
* 29568 29667: gap of 100 bp
*
* 29668 40520: contig of 10853 bp in length
*
* 40521 40620: gap of 100 bp
*
* 40621 50002: contig of 9382 bp in length
*
* 50003 50102: gap of 100 bp
*
* 50103 54942: contig of 4840 bp in length
*
* 54943 55042: gap of 100 bp
*
* 55043 59067: contig of 4025 bp in length
*
* 59068 59167: gap of 100 bp
*
* 59168 72327: contig of 13160 bp in length
*
* 72328 72427: gap of 100 bp
*
* 72428 79396: contig of 6969 bp in length
*
* 79397 79496: gap of 100 bp
*
* 79497 82738: contig of 3242 bp in length
*
* 82739 82838: gap of 100 bp
*
* 82839 87543: contig of 4705 bp in length
*
* 87544 87643: gap of 100 bp
*
* 87644 89847: contig of 2204 bp in length
*
* 89848 89947: gap of 100 bp
*
* 89948 92376: contig of 2429 bp in length
*
* 92377 92476: gap of 100 bp
*
* 92477 99799: contig of 7323 bp in length
*
* 99800 99899: gap of 100 bp
*
* 99900 102117: contig of 2218 bp in length
*
* 102118 102217: gap of 100 bp
*
* 102218 109330: contig of 7113 bp in length
*
* 109331 109430: gap of 100 bp
*
* 109431 112187: contig of 2757 bp in length
*
* 112188 112287: gap of 100 bp
*
* 112288 116407: contig of 4120 bp in length
*
* 116408 116507: gap of 100 bp
*
* 116508 119386: contig of 2879 bp in length
*
* 119387 119486: gap of 100 bp
*
* 119487 121869: contig of 2383 bp in length
*
* 121870 121969: gap of 100 bp
*
* 121970 125546: contig of 3577 bp in length
*
* 125547 125646: gap of 100 bp
*
* 125647 128990: contig of 3344 bp in length
*
* 128991 129090: gap of 100 bp
*
* 129091 134055: contig of 4965 bp in length
*
* 134056 134155: gap of 100 bp
*
* 134156 138314: contig of 4159 bp in length
*
* 138315 138414: gap of 100 bp
*
* 138415 140612: contig of 2198 bp in length
*
* 140613 140612: contig of 2198 bp in length
```

```

* 140613 140712: gap of 100 bp
* 140713 143296: contig of 2584 bp in length
* 143297 143396: gap of 100 bp
* 143397 148113: contig of 4717 bp in length
* 148114 148213: gap of 100 bp
* 148214 151009: contig of 2796 bp in length
* 151010 151109: gap of 100 bp
* 151110 156616: contig of 5507 bp in length
* 156617 156716: gap of 100 bp
* 156717 158763: contig of 2047 bp in length
* 158764 158863: gap of 100 bp
* 158864 162129: contig of 3266 bp in length
* 162130 162229: gap of 100 bp
* 162230 164443: contig of 2214 bp in length
* 164444 164543: gap of 100 bp
* 164544 169800: contig of 5257 bp in length
* 169801 169900: gap of 100 bp
* 169901 251545: contig of 81645 bp in length
* 251546 251645: gap of 100 bp
* 251646 259029: contig of 7384 bp in length
* 259030 259129: gap of 100 bp
* 259130 262871: contig of 3742 bp in length
* 262872 262971: gap of 100 bp
* 262972 265209: contig of 2238 bp in length
* 265210 265309: gap of 100 bp
* 265310 267581: contig of 2272 bp in length.

```

FEATURES

source

```

1. .267581
/organism="Homo sapiens"
/db_xref="taxon:9606"
/Chromosome="9"
/clone="RP11-133022"
/clone_lib="RPC1-11.1"
1. .25718
/note="assembly_fragment:00645"
fragment_chain:1
clone_end:T7
vector_side:left
25819..29567
/note="assembly_fragment:01660"
fragment_chain:1
29668..40520
/note="assembly_fragment:03080"
fragment_chain:1
40621..50002
/note="assembly_fragment:00771"
fragment_chain:1
50103..54942
/note="assembly_fragment:02629"
fragment_chain:1
55043..59067
/note="assembly_fragment:04565"
fragment_chain:1
59168..72327
/note="assembly_fragment:00223"
fragment_chain:1
72428..79396
/note="assembly_fragment:03318"
fragment_chain:1
79497..82738
/note="assembly_fragment:00224"
fragment_chain:2
82839..87543
/note="assembly_fragment:03050"
fragment_chain:2
87644..89847
/note="assembly_fragment:01948"
fragment_chain:3
89948..92376
/note="assembly_fragment:02191"
fragment_chain:3
92477..99799
/note="assembly_fragment:00010"

```

```

misc_feature
99900..102117
/note="assembly_fragment:00672"
102218..109330
/note="assembly_fragment:01326"
109431..112187
/note="assembly_fragment:01498"
112288..116407
/note="assembly_fragment:01607"
116508..119386
/note="assembly_fragment:01899"
119487..121869
/note="assembly_fragment:02068"
121970..125546
/note="assembly_fragment:02469"
125647..128990
/note="assembly_fragment:02502"
129091..134035
/note="assembly_fragment:02669"
134156..138314
/note="assembly_fragment:02757"
138415..140612
/note="assembly_fragment:02776"
140713..143296
/note="assembly_fragment:03127"
143397..148113
/note="assembly_fragment:03634"
148214..151009
/note="assembly_fragment:03898"
151110..156616
/note="assembly_fragment:04220"
156717..158763
/note="assembly_fragment:04276"
158864..162129
/note="assembly_fragment:04418"
162230..164443
/note="assembly_fragment:04428"

```

```

Query Match      64.7%; Score 1317.8; DB 80; Length 267581;
Best Local Similarity 91.9%; Pred. No. 0;
Matches 1355; Conservative 4; Mismatches 113; Indels 3; Gaps 3;

```

```

QY 442 cagagatccctgggagaaatgccggccgccatcttgggtcactgatgcctgcgcctg 501
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165900 CACAGATCCCTGGGAGAAATGCCGGCCGCCATCTTTGGGTCTATCGATGAGCTCGCCCTG 165841

QY 502 tcctgtgtcccttctgaggaagacattagaaataaattgattgtcttcttaag 561
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165840 TGCCTGTCTCCGCTTGTGAGGGAAGACATAGAAATGAATGTGTCTCTTAAGG 165781

QY 562 atgggcaggaacacagatcctgtgtggatatttattgaacggggtacagatttga 621
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165780 ATGGGCAGGAAACACAGATCCTGTTGTGGATATTTATTGAACGGGATTACAGATTTGAAA 165721

QY 622 tgaagtcaaaagttagaattaccatgagagagaaacagagagaaatcttgatgct 681
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165720 TGAAGTCACAAGTGTAGCATTACCAATGAGAGGAAACACAGAGAAAAATCTTTGATGGCT 165661

QY 682 tcacaagacatcaacaacaaatggaatactgtgatgaatgagcagcagcaagctgg 741
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165660 TCACAAGACATGCAACAAACAAATGGAATACGTGTGATGACATGAGCGACCAAGCTGGG 165601

QY 742 gaggagataaccacggggcagaggggtcaggattcttgccctgctgcctaaactgtcg 801
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165600 GAGGAGATAACACACGGGCGAGAGGGTTCAGGATTTCTGCCCTGCTGCTAACTGTCCGTT 165541

QY 802 cataaccaaatatttcatttcttaacctcaaaacaaagctgttgcataatcctgatct 861
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165540 CATACACAAATCATTTTCATATTTCTAACCCCTCAAAACAAAGCTGTTGTATATCTGATCT 165481

QY 862 ctacggttctctggcccaacattctccatatatccagccacacatcttttaatt 921
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165480 CTACGGTTCCTTCTGGGCCCCAACATTTCTCATATATATCCAGCCCACTCAATTTTAAATAT 165421

```


Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 138647 bases at least Q40
Consensus quality: 140619 bases at least Q30
Consensus quality: 141706 bases at least Q20
Insert size: 142775; sum-of-contigs
Insert size: 147523; 5.9% error; agarose-fp
Quality coverage: 4.29x in Q20 bases; sum-of-contigs Quality
coverage: 4.22x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 12163: contig of 12163 bp in length
* 12164 12263: gap of 100 bp
* 12264 25269: contig of 13006 bp in length
* 25270 25369: gap of 100 bp
* 25370 57709: contig of 32340 bp in length
* 57710 57809: gap of 100 bp
* 57810 66792: contig of 8983 bp in length
* 66793 66922: gap of 100 bp
* 66923 102772: contig of 35880 bp in length
* 102773 102872: gap of 100 bp
* 102873 106863: contig of 3991 bp in length
* 106864 106963: gap of 100 bp
* 106964 129804: contig of 22841 bp in length
* 129805 129904: gap of 100 bp
* 129905 132223: contig of 2319 bp in length
* 132224 132323: gap of 100 bp
* 132324 135764: contig of 3441 bp in length
* 135765 135864: gap of 100 bp
* 135865 143675: contig of 7811 bp in length.

FEATURES

source
1. .143675
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-146P9"
/clone_lib="RPC1-11.1"
/note="assembly_fragment:00440
vector_side:left"
12264..25269
/note="assembly_fragment:00434
fragment_chain:1"
25370..57709
/note="assembly_fragment:00526
fragment_chain:1"
57810..66792
/note="assembly_fragment:00661
fragment_chain:1"
66893..102772
/note="assembly_fragment:00323"
102873..106863
/note="assembly_fragment:00860"
106964..129804
/note="assembly_fragment:00978"
129905..132223
/note="assembly_fragment:01068"
132324..135764
/note="assembly_fragment:01424"
135865..143675
/note="assembly_fragment:01413
clone_end:sp6
vector_side:right"
42329 a 30042 c 28964 g 41431 t 909 others

BASE COUNT 42329 a 30042 c 28964 g 41431 t 909 others
ORIGIN

Query Match 4.9%; Score 99; DB 79; Length 143675;
Best Local Similarity 100.0%; Pred. No. 8.9e-14;
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaagctgcatacagaacacagagggagattgtgtgctgcagccgagggagaccag 60
|||||
Db 136065 AGAAGCTGCATCAGAAAAACAGAGGGGAGATTGTGTGCTGCAGCCGAGGAGACCAG 136124

QY 61 gaagatctcatggtgggaagacacctgatgatacagagg 99
|||||
Db 136125 GAAGATCTCATGTTGGTGGGAAGGACCTGATGATACAGAGG 136163

RESULT 8

AL358573

LOCUS

AL358573 267581 bp DNA HTG 15-APR-2001

PROGRESS ***; 37 unordered pieces.

ACCESSION

AL358573

VERSION

AL358573.17 GI:13660951

KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 267581)

AUTHORS

Mclay, K.

TITLE

Direct Submission

Submitted (14-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Apr 17, 2001 this sequence version replaced gi:13398774.

COMMENT

----- Genome Center

Center: Sanger Centre

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: bA133022

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 256699 bases at least Q40

Consensus quality: 259744 bases at least Q30

Consensus quality: 261407 bases at least Q20

Insert size: 263981; sum-of-contigs

Insert size: 135491; 19.3% error; agarose-fp

Quality coverage: 5.51x in Q20 bases; sum-of-contigs Quality

coverage: 12.23x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently

* consists of 37 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 25718: contig of 25718 bp in length

* 25719 25818: gap of 100 bp

* 25819 29567: contig of 3749 bp in length

* 29568 29667: gap of 100 bp

* 29668 40520: contig of 10853 bp in length

* 40521 40620: gap of 100 bp

* 40621 50002: contig of 9382 bp in length

* 50003 50102: gap of 100 bp

* 50103 54942: contig of 4840 bp in length

* 54943 55042: gap of 100 bp

* 55043 59067: contig of 4025 bp in length

* 59068 59167: gap of 100 bp

* 59168 72327: contig of 13160 bp in length

* *


```

* 72328 72427: gap of 100 bp
* 72428 79396: contig of 6969 bp in length
* 79397 79496: gap of 100 bp
* 79497 82738: contig of 3242 bp in length
* 82739 82838: gap of 100 bp
* 82839 87543: contig of 4705 bp in length
* 87544 87643: gap of 100 bp
* 87644 89847: contig of 2204 bp in length
* 89848 89947: gap of 100 bp
* 89948 92376: contig of 2429 bp in length
* 92377 92476: gap of 100 bp
* 92477 99799: contig of 7323 bp in length
* 99800 99899: gap of 100 bp
* 99900 102117: contig of 2218 bp in length
* 102118 102217: gap of 100 bp
* 102218 109330: contig of 7113 bp in length
* 109331 109430: gap of 100 bp
* 109431 112187: contig of 2757 bp in length
* 112188 112287: gap of 100 bp
* 112288 116407: contig of 4120 bp in length
* 116408 116507: gap of 100 bp
* 116508 119386: contig of 2879 bp in length
* 119387 119486: gap of 100 bp
* 119487 121869: contig of 2383 bp in length
* 121870 121969: gap of 100 bp
* 121970 125546: contig of 3577 bp in length
* 125547 125646: gap of 100 bp
* 125647 128990: contig of 3344 bp in length
* 128991 129090: gap of 100 bp
* 129091 134055: contig of 4965 bp in length
* 134056 134155: gap of 100 bp
* 134156 138314: contig of 4159 bp in length
* 138315 138414: gap of 100 bp
* 138415 140612: contig of 2198 bp in length
* 140613 140712: gap of 100 bp
* 140713 143296: contig of 2584 bp in length
* 143297 143396: gap of 100 bp
* 143397 148113: contig of 4717 bp in length
* 148114 148213: gap of 100 bp
* 148214 151009: contig of 2796 bp in length
* 151010 151109: gap of 100 bp
* 151110 156616: contig of 5507 bp in length
* 156617 156716: gap of 100 bp
* 156717 158763: contig of 2047 bp in length
* 158764 158863: gap of 100 bp
* 158864 162129: contig of 3266 bp in length
* 162130 162229: gap of 100 bp
* 162230 164443: contig of 2214 bp in length
* 164444 164543: gap of 100 bp
* 164544 169800: contig of 5257 bp in length
* 169801 169900: gap of 100 bp
* 169901 251545: contig of 81645 bp in length
* 251546 251645: gap of 100 bp
* 251646 259029: contig of 7384 bp in length
* 259030 259129: gap of 100 bp
* 259130 262871: contig of 3742 bp in length
* 262872 262971: gap of 100 bp
* 262972 265209: contig of 2238 bp in length
* 265210 265309: gap of 100 bp
* 265310 267581: contig of 2272 bp in length.
FEATURES
    source
        1..267581
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="9"
            /clone="RP11-133022"
            /clone.lib="RPC1-11.1"
            1..25718
                /note="assembly_fragment:00645
                fragment_chain:1
                clone_end:T7
                vector_side:left"
            25819..29567
                misc_feature
                    /note="assembly_fragment:00645
                    fragment_chain:1
                    misc_feature
                        /note="assembly_fragment:00645
                        fragment_chain:1

```

```

/misc_feature
    /note="assembly_fragment:01660
    fragment_chain:1"
    29668..40520
        /note="assembly_fragment:03080
        fragment_chain:1"
        40621..50002
            /note="assembly_fragment:00771
            fragment_chain:1"
            50103..54942
                /note="assembly_fragment:02629
                fragment_chain:1"
                55043..59067
                    /note="assembly_fragment:04565
                    fragment_chain:1"
                    59168..72327
                        /note="assembly_fragment:00223
                        fragment_chain:1"
                        72428..79396
                            /note="assembly_fragment:03318
                            fragment_chain:1"
                            79497..82738
                                /note="assembly_fragment:00224
                                fragment_chain:2"
                                82839..87543
                                    /note="assembly_fragment:03050
                                    fragment_chain:2"
                                    87644..89847
                                        /note="assembly_fragment:01948
                                        fragment_chain:3"
                                        89948..92376
                                            /note="assembly_fragment:02191
                                            fragment_chain:3"
                                            92477..99799
                                                /note="assembly_fragment:00010"
                                                99900..102117
                                                    /note="assembly_fragment:00672"
                                                    102218..109330
                                                        /note="assembly_fragment:01326"
                                                        109431..112187
                                                            /note="assembly_fragment:01498"
                                                            112288..116407
                                                                /note="assembly_fragment:01607"
                                                                116508..119386
                                                                    /note="assembly_fragment:01899"
                                                                    119487..121869
                                                                        /note="assembly_fragment:02068"
                                                                        121970..125546
                                                                            /note="assembly_fragment:02469"
                                                                            125647..128990
                                                                                /note="assembly_fragment:02502"
                                                                                129091..134055
                                                                                    /note="assembly_fragment:02669"
                                                                                    134156..138314
                                                                                        /note="assembly_fragment:02757"
                                                                                        138415..140612
                                                                                            /note="assembly_fragment:02776"
                                                                                            140713..143296
                                                                                                /note="assembly_fragment:03127"
                                                                                                143397..148113
                                                                                                    /note="assembly_fragment:03634"
                                                                                                    148214..151009
                                                                                                        /note="assembly_fragment:03898"
                                                                                                        151110..156616
                                                                                                            /note="assembly_fragment:04220"
                                                                                                            156717..158763
                                                                                                                /note="assembly_fragment:04276"
                                                                                                                158864..162129
                                                                                                                    /note="assembly_fragment:04418"
                                                                                                                    162230..164443
                                                                                                                        /note="assembly_fragment:04428"

```

Query Match 4.9%; Score 99; DB 80; Length 267581;
 Best Local Similarity 100.0%; Pred. No. 9.4e-14;

```
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 agaagctgcatcagaaaaacagaggagattgtgtgctgcagccgaggagaccag 60
|||||
Db 253275 AGAAGCTGCATCAGAAAACAGAGGGAGATTGTGTGGCTGCACCCGAGGAGACCAG 253334
QY 61 gaagatctgcattgtgtgggaagacctgatgatacagagg 99
|||||
Db 253335 GAAGATCTGCATGTTGGGAAGGACCTGATGATACAGAGG 253373

RESULT 9
AF279290 580 bp DNA PRI 04-DEC-2000
LOCUS Homo sapiens prostate-cancer-specific DD3 protein gene, promoter
DEFINITION AF279290
ACCESSION AF279290
VERSION AF279290.1 GI:11528086
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 580)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Verhaegh,G.W., van Bokhoven,A., Smit,F., Schalken,J.A. and
Bussemakers,M.J.G.
Isolation and Characterization of the Promoter of the Human
Prostate Cancer-specific DD3 Gene
J. Biol. Chem. 275 (48), 37496-37503 (2000)
PUBMED 10982808
TITLE 2 (bases 1 to 580)
AUTHORS Verhaegh,G.W., van Bokhoven,A., Smit,F., Schalken,J.A. and
Bussemakers,M.J.G.
Direct Submission
Submitted (16-JUN-2000) Urology Research Laboratory, University
Medical Center Nijmegen, Geert Grooteplein Zuid 10, Nijmegen 6525
GA, The Netherlands
FEATURES
source
1..580
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/map="9q21-q22"
1..460
promoter 260..288
protein_bind /bound_moiety="high mobility group I Y protein"
461..580
mRNA /product="prostate-cancer-specific DD3 protein"
461..580
exon /number=1
BASE COUNT 190 a 97 c 153 g 140 t
ORIGIN
Query Match 4.8%; Score 98; DB 89; Length 580;
Best Local Similarity 100.0%; Pred. No. 9.5e-14;
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 agaagctgcatcagaaaaacagaggagattgtgtgctgcagccgaggagaccag 60
Db 483 AGAAGCTGCATCAGAAAACAGAGGGAGATTGTGTGGCTGCACCCGAGGAGACCAG 542
QY 61 gaagatctgcattgtgtgggaagacctgatgatacagag 98
|||||
Db 543 GAAGATCTGCATGTTGGGAAGGACCTGATGATACAGAG 580

RESULT 10
AC009556/c
LOCUS Homo sapiens clone RP11-57C21, WORKING DRAFT SEQUENCE, 26 unordered
DEFINITION pieces
ACCESSION AC009556
AC009556.4 GI:7107752
HTGS_PHASE1; HTGS_DRAFT.
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome, clone RP11-57C21
Unpublished
2 (bases 1 to 172298)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Collangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArelano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferraira,P., Fitzhugh,W., Forrest,C.,
Funke,B., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 28, 2000 this sequence version replaced gi:479158.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1672
Center clone name: 57_C21
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-primer-amersham; 5% of reads
Chemistry: Dye-terminator Big Dye; 95% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 136385 bases at least Q40
Consensus quality: 156343 bases at least Q30
Consensus quality: 165534 bases at least Q20
Insert size: 165000; agarose-fp
Insert size: 169798; sum-of-contigs
Quality coverage: 4.2 in Q20 bases; agarose-fp
Quality coverage: 4.1 in Q20 base.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 1016: contig of 1016 bp in length
* 1017 1116: gap of 100 bp
* 1117 2136: contig of 1020 bp in length
* 2137 2236: gap of 100 bp
* 2237 3571: contig of 1335 bp in length
* 3572 3671: gap of 100 bp
* 3672 5272: contig of 1601 bp in length
* 5273 5372: gap of 100 bp
* 5373 6381: contig of 1009 bp in length
* 6382 6481: gap of 100 bp
* 6482 8508: contig of 2027 bp in length
* 8509 8608: gap of 100 bp
```


[illegible]

	Query Match	2.6%	Score 52.6;	DB 10;	Length 1141;
	Best Local Similarity	9.2%;	Pred. NO. 0.019;		
	Matches	68;	Conservative 282;	Mismatches 385;	Indels 1; Gaps 1;
Qy	1162	tttgccatactgaaattcatttcccaattttgtgcccattctcaagaccctcaaaatgctc	1221		
		::: :::: :: :: :::: :::: :::: :::: :::: :::: :::: :::: ::::			
Db	165	WKKNNNATGWD DDTKYHMMNNNGCBTVYMWRYKTRDWSBKRNMYGMBWKNWSYDVT	224		
Qy	1222	attccattaaatcacagagattaaacttttttttaacctggaagaattcaattcttagcat	1281		
		::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::			
Db	225	YYWVWDDMCRRKVRWRVTRGRNRYVAVBETAHRRRYNNGWTBAMAYRRWTMNNNNN	284		
Qy	1282	gcagctaaggaaatttaattacataattttgttttccagtcgaaagatgactaaagtccttt	1341		
		:::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::			
Db	285	AKAMCKRAKYGWGNRABVYNSTCTWKSSTTKVTRTSWANNCRAGDANKDHHKWWSAAMG	344		
Qy	1342	atccctcccccttttttttgaatttttttttccactataaagttaaaatgcttaagcctttgatact	1401		

Db	285	AKAMCKRAKYWGNRRABVNSTCTTKWSKTTKVRTSWANNRCAGDANKDHKWKWSAAMG	344
Qy	1342	atccctcccccttggttgattttttccagataaaagtctagcctgtact	1401
Db	345	VYNNNNNNNTYTKARHBARDWVHWSAWKWHAANAHHYSRKKWTBYKRKTMVNNNGT	404
Qy	1402	gaagctgtatacagacagacccctcccccatccctccagccttatctgtcataccacataa	1461
Db	405	TMKRKMAWYTKMDMDWBGTYNNNNGGRTYGGWTKNKMTYYKWKANNCKRWADHKT	464
Qy	1462	ccccctcccatnysaacataacaaaatctaacttgtaattctctgaacatgtcagncata	1521
Db	465	CTHNNTTWKMKTYWNCYKSGTNGKSHRBAAAYTYWYMWRRYAHANNNNNDYWKKA	524
Qy	1522	cahttctctctgcctgagaagctctctctctcttaantctagaatgatataagt	1581
Db	525	CTWYKVBVCSKWNWNYAAYTYKSSNNYTSRYTRWNTNSWRWSDTRSGRNNYARABH	584
Qy	1582	tttgaataagttgactatcttctcactgaagaaggacacatagatgattcatcatc	1641
Db	585	YGYKWTNRWBWSHTWBHBRAAAHYWBMYYBAKCHCKMAWYRAKKYAGAGSGNNNNN	644
Qy	1642	acatgagacgcgaatacataaagtgttaatttgattataagagtttagataataatga	1701
Db	645	NNNNNNNNNATCAREDYAAASRWYAMANKWYVYKBAANNAYTYHANNWGWGNNATDT	704

* 70805 73600: contig of 2796 bp in length
 * 73601 73700: gap of unknown length
 * 73701 75760: contig of 2060 bp in length
 * 75761 75860: gap of unknown length
 * 75861 78368: contig of 2508 bp in length
 * 78369 78468: gap of unknown length
 * 78469 81134: contig of 2666 bp in length
 * 81134 81234: gap of unknown length
 * 81235 82918: contig of 1584 bp in length
 * 82919 85499: contig of 2581 bp in length
 * 85500 85599: gap of unknown length
 * 85600 87735: contig of 2136 bp in length
 * 87736 87835: gap of unknown length
 * 87836 90615: contig of 2780 bp in length
 * 90616 93145: contig of 2430 bp in length
 * 93146 93246: gap of unknown length
 * 93246 94702: contig of 1457 bp in length
 * 94703 94802: gap of unknown length
 * 94803 96523: contig of 1721 bp in length
 * 96524 96624: gap of unknown length
 * 96624 98319: contig of 1696 bp in length
 * 98320 100330: contig of 1811 bp in length
 * 100331 100330: gap of unknown length
 * 100331 101717: contig of 1387 bp in length
 * 101718 101817: gap of unknown length
 * 101818 103330: contig of 1513 bp in length
 * 103331 103430: gap of unknown length
 * 103431 104873: contig of 1443 bp in length
 * 104874 104973: gap of unknown length
 * 104974 107045: contig of 2072 bp in length
 * 107046 107145: gap of unknown length
 * 107146 108189: contig of 1044 bp in length
 * 108190 108289: gap of unknown length
 * 108290 109896: contig of 1407 bp in length
 * 109897 117099: contig of 1913 bp in length
 * 117100 118109: gap of unknown length
 * 118110 118819: contig of 2010 bp in length
 * 118820 119325: contig of 1538 bp in length
 * 119326 119324: gap of unknown length
 * 119325 119324: gap of unknown length
 * 119325 120634: contig of 1310 bp in length
 * 120635 120734: gap of unknown length
 * 120735 121847: contig of 1113 bp in length
 * 121848 121947: gap of unknown length
 * 121948 123179: contig of 1232 bp in length
 * 123180 123279: gap of unknown length
 * 123280 125035: contig of 1756 bp in length
 * 125036 125135: gap of unknown length
 * 125136 126204: contig of 1059 bp in length
 * 126205 126304: gap of unknown length
 * 126305 127411: contig of 1107 bp in length
 * 127412 127511: gap of unknown length
 * 127512 129229: contig of 1718 bp in length
 * 129230 129329: gap of unknown length
 * 129330 130547: contig of 1218 bp in length
 * 130548 130647: gap of unknown length
 * 130648 131798: contig of 1151 bp in length
 * 131799 131898: gap of unknown length
 * 131899 133127: contig of 1229 bp in length
 * 133128 133227: gap of unknown length
 * 133228 134591: contig of 1364 bp in length
 * 134592 134691: gap of unknown length
 * 134692 135736: contig of 1045 bp in length
 * 135737 135836: gap of unknown length
 * 135837 137064: contig of 1228 bp in length.

Query Match 2.3%; Score 47.2; DB 74; Length 137064;
 Best Local Similarity 53.2%; Pred. No. 0.65;
 Matches 100; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
 QY 1181 atttcccaattttgtccattctcaagacctcaaaatgtcattccattcaataatcagcag 1240
 Db 107698 ATCTCTACTAGGAATACCTTCCTATTCGTTGTCTATTAAGTCAATTAATGCTGA 107757
 QY 1241 attaaatttttttttaacctggaagaattcaatgttacatgcagctatgggaattaat 1300
 Db 107758 AGTAAATTTATGTTTAAATTTTATGCTTAAGTTGGATATATCTGTAAGGGAATTTACT 107817
 QY 1301 tacataattttgtttccagtgcaagatgactaagtccttattccctccctttgtttga 1360
 Db 107818 TGAGTTTTTTTTTTTTTTCATAAAGCTGCCCTAATGTTTACCACAATATTTAAATAA 107877
 QY 1361 tttttttt 1368
 Db 107878 ACTATCTT 107885

RESULT 15
 AC021589/c
 LOCUS Homo sapiens clone RP11-140G18, WORKING DRAFT SEQUENCE, 16
 DEFINITION unordered pieces.
 ACCESSION AC021589
 VERSION AC021589.3 GI:7596811
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 152571)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens, clone RP11-140G18
 2 (bases 1 to 152571)
 Unpublished
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
 Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
 Choepell,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
 Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
 Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
 Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
 Landers,T., Lechoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
 Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
 McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
 Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,
 Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
 Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
 Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
 Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 19, 2000 this sequence version replaced gi:7331450.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L5774
 Center clone name: 140_G18
 ----- Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads

TITLE
 JOURNAL
 COMMENT

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 16:50:00 ; Search time 46.06 Seconds
(without alignments)
146.495 Million cell updates/sec

Title: US-09-402-713A-2

Perfect score: 268

Sequence: 1 MFLHSSPPKYPHTQEAQKE.....HLGSSMSLALCLVPLVREGH 51

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL16:*

- 1: sp.archaea:*
- 2: sp.bacteria:*
- 3: sp.fungi:*
- 4: sp.human:*
- 5: sp.invertebrate:*
- 6: sp.mammal:*
- 7: sp.mhc:*
- 8: sp.organelle:*
- 9: sp.phase:*
- 10: sp.plant:*
- 11: sp.rodent:*
- 12: sp.unclassified:*
- 13: sp.vertebrate:*
- 14: sp.virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	60.5	22.6	1236	5 Q9NKF9	Q9nkf9 drosophila
2	60.5	22.6	1238	5 Q9VJW9	Q9vjw9 drosophila
3	60.5	22.6	1239	5 Q94902	Q94902 drosophila
4	60	22.4	187	2 Q9JUF3	Q9juf3 neisseria m
5	60	22.4	218	2 Q9JRV5	Q9jrv5 neisseria m
6	59	22.0	1114	11 Q9JL73	Q9jlv3 mus musculus
7	56.5	21.1	1061	2 P73998	P73998 synechocyst
8	56.5	21.1	6396	2 Q9KID7	Q9kid7 streptomyces
9	56	20.9	338	2 Q9RVU4	Q9rvu4 deinococcus
10	56	20.9	433	5 Q9V7N7	Q9v7n7 drosophila
11	56	20.9	558	5 Q9NFP1	Q9nfp1 drosophila
12	55.5	20.7	698	2 Q9FLJ5	Q9flj5 pseudomonas
13	55	20.5	168	14 Q9E138	Q9e138 bovine vira
14	55	20.5	196	5 Q9VKF5	Q9vkf5 drosophila
15	55	20.5	469	5 Q9GRT5	Q9grt5 leishmania
16	55	20.5	1032	11 Q61989	Q61989 mus musculus
17	55	20.5	1935	5 Q9VQW0	Q9vqw0 drosophila
18	54.5	20.3	116	14 Q92886	Q92886 human immun
19	54.5	20.3	177	6 Q9XS95	Q9xs95 oryctolagus

20	54.5	20.3	244	2 Q9HYW1	Q9hyw1 pseudomonas
21	54.5	20.3	305	10 Q9MIC3	Q9mic3 arabisdopsis
22	54.5	20.3	461	11 Q9R1V8	Q9rlv8 mus musculus
23	54	20.1	204	3 Q9P3L8	Q9p3l8 neurospora
24	54	20.1	257	2 Q9Z8E3	Q9z8e3 chlamydia p
25	54	20.1	450	10 Q9FMN6	Q9fmn6 arabisdopsis
26	54	20.1	633	2 Q9HTU6	Q9htu6 pseudomonas
27	54	20.1	784	4 Q00146	Q00146 homo sapien
28	53.5	20.0	250	8 Q33122	Q33122 schismocarp
29	53.5	20.0	348	8 Q35379	Q35379 paracyprich
30	53.5	20.0	408	4 Q9UGL4	Q9ugl4 homo sapien
31	53.5	20.0	421	8 Q9TIC2	Q9tic2 balbiana i
32	53.5	20.0	443	8 Q32878	Q32878 palmaria pa
33	53.5	20.0	449	2 Q9Z9E0	Q9z9e0 chlamydia p
34	53.5	20.0	783	4 Q9NXD0	Q9nxd0 homo sapien
35	53	19.8	671	5 Q9W365	Q9w365 drosophila
36	53	19.8	1161	11 Q9QYE7	Q9qye7 rattus norv
37	53	19.8	2205	5 Q9NGQ2	Q9ngq2 dictyosteli
38	52.5	19.6	216	2 Q9RVK0	Q9rvk0 deinococcus
39	52.5	19.6	357	2 Q9I720	Q9i720 pseudomonas
40	52.5	19.6	360	4 Q9H9N9	Q9h9n9 homo sapien
41	52.5	19.6	384	8 Q32401	Q32401 halosaccion
42	52.5	19.6	392	8 Q31855	Q31855 audouinella
43	52.5	19.6	424	8 Q9MV15	Q9mv15 ballia call
44	52.5	19.6	488	8 Q9THF8	Q9thf8 palmaria pa
45	52.5	19.6	493	4 Q9UF35	Q9uf35 homo sapien

ALIGNMENTS

RESULT 1

Q9NKF9 PRELIMINARY; PRT; 1236 AA.
AC Q9NKF9;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE KUZ PROTEIN.
GN KUZ OR BG:DS07660.3 OR CG7147.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y, CN BW SP;
RX MEDLINE=99403001; PubMed=10471707;
RA Ashburner M., Misra S., Roote J., Lewis S.E., Blazej R., Davis T.,
RA Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,
RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
RA Palazziolo M., Reese M.G., Spradling A., Tsang C., Wan K., Whitelaw K.,
RA Ceiniker S., Rubin G.M.;
RT "An exploration of the sequence of a 2.9-Mb region of the genome of
RT Drosophila melanogaster: the Adh region.";
RL Genetics 153:179-219(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Y, CN BW SP;
RA Ceiniker S.E., Aqbayani A., Arcaina T.T., Baxter E., Blazej R.G.,
RA Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
RA Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
RA Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,
RA Lewis S., Li P., Lomtan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
RA Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,
RA Sethi H., Snir E., Swirskas R.R., Wan K.H., Weinburg T., Zhang R.,
RA Zieran L.L., Rubin G.M.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003407; AAF4800.1; -;
DR FlyBase; FBgn0015954; kuz.
DR InterPro; IPR000130; -;
DR InterPro; IPR001590; -;

```
DR InterPro: IPR001762; -.
DR PROSITE: PS0215; ADAM_MEPRO; 1.
DR PROSITE: PS0214; DISINTEGRIN_2; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR SMART: SM00050; DISIN; 1.
SQ SEQUENCE 1236 AA; 136343 MW; 40461AC17040C9AD CRC64;

Query Match 22.6%; Score 60.5; DB 5; Length 1236;
Best Local Similarity 47.2%; Pred. No. 19;
Matches 17; Conservative 4; Mismatches 10; Indels 5; Gaps 2;

QY 9 FKYPHTQEAQKQAQRSLGEM----PGRHLGSSMSLA 40
Db 338 FKYPH-QKYTKANFAEGAFYDPSTGRRLLGSSANVA 372
||||| : ||| : | ||||| :|

RESULT 2
Q9VJW9 PRELIMINARY; PRT; 1238 AA.
AC Q9VJW9;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE KUZ PROTEIN.
GN KUZ OR BG:DS07660.3 OR CG7147.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananidis P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Poilard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shie B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner C., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
```



```

Query Match      20.9%; Score 56; DB 5; Length 558;
Best local Similarity 33.3%; Pred. No. 32;
Matches 12; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

Qy 3 LHISSPKYHPHTQAQKEAQRSLGELGEMPCRHILGSSMS 38
    ::::  |||  :  :  :  |||  |||  :  :  :  |||  :  :  :  |||  :
Db 292 IYVAQVQYPTTSRLSKKKRAKRSLOEIQTRQVASSSS 327

```

RESULT	14				
Q9VKF5					
ID	Q9VKF5	PRELIMINARY;	PRT;	196 AA.	
AC	Q9VKF5;				
DT	01-MAY-2000	(TrEMBLrel. 13, Created)			

```

Query Match      20.9%; Score 56; DB 5; Length 558;
Best Local Similarity 33.3%; Pred. No. 32;
Matches 12; Conservative 10; Mismatches 14; Indels

```

DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE CG14931 PROTEIN.
GN CG14931.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
ON NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Adayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Merkulov G., Milshina N.V., Mobarra C., McLeod M.P., McPherson D.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AF003633; AAF53117.1;
DR FlyBase; FBgn0032374; CG14931.
SQ SEQUENCE 196 AA; 21307 MW; 5E90E9EF16EEF2B4 CRC64;

Query Match 20.5%; Score 55; DB 5; Length 196;
Best Local Similarity 61.1%; Pred. No. 14;
Matches 11; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 24 SGENPGRHLGSSMSLAL 41
|:||||| ||||| :|:|:
Db 167 SIGEMPDRLHISTSSVAV 184

RESULT 15

Q9GRT5 PRELIMINARY; PRT; 469 AA.
AC Q9GRT5;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE CYCLIN-DEPENDENT KINASE.

GN CRK.
OS Leishmania mexicana.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MNYC/Bz/62/M379;
RA Wiese M.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MNYC/Bz/62/M379;
RA Martin W., Wang Q., Goercke I.;
RT "Identification of mitogen-activated protein (MAP) kinases from
RT Leishmania mexicana."
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ293288; CAC07964.1; -.
KW Kinase.
SQ SEQUENCE 469 AA; 52184 MW; C17DFD515BE2FD85 CRC64;

Query Match 20.5%; Score 55; DB 5; Length 469;
Best Local Similarity 29.8%; Pred. No. 36;
Matches 14; Conservative 7; Mismatches 18; Indels 8; Gaps 1;

QY 1 MELHISSPEK-----YPHTQEAQKQKRSIGEMPGRRHLGSSMSL 39
||:||||| ||||| :|:|:
Db 287 MEWLLGSPTRETWPSMPSYPTGIMERLERASRAAERPDLRFSGSDVTV 333

Search completed: July 30, 2001, 16:53:06
Job time: 186 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 16:50:20 ; Search time 17.83 Seconds
(without alignments)

97.983 Million cell updates/sec

Title: US-09-402-713a-2.

Perfect score: 268

Sequence: 1 MELHISSPKPYHTQEAQK.....HLGSSMSLALCLVPLVREGH 51

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	60.5	22.6	343	1	TA2R_BOVIN
2	59	22.0	1113	1	HDA5_MOUSE
3	59	22.0	1122	1	HDA5_HUMAN
4	55	20.5	1039	1	ITPA_MOUSE
5	54.5	20.3	461	1	CO1A_BOVIN
6	54.5	20.3	461	1	CO1A_HUMAN
7	54.5	20.3	461	1	CO1A_MOUSE
8	53.5	20.0	446	1	GNTU_ECOLI
9	53.5	20.0	451	1	PTCC_BACST
10	53.5	20.0	580	1	5NTD_BOOMI
11	53	19.8	554	1	DCS4_GOSAR
12	52.5	19.6	441	1	XRRA_BRARE
13	52	19.4	337	1	RIR2_TRYBB
14	52	19.4	2774	1	MAPA_RAT
15	51.5	19.2	269	1	AQPI_HOMO
16	51.5	19.2	470	1	RBL_PROHO
17	51.5	19.2	569	1	4CL2_ORYSA
18	51	19.0	352	1	KEA_BRARE
19	51	19.0	1038	1	ITA4_HUMAN
20	51	19.0	1162	1	ITAD_HUMAN
21	51	19.0	2485	1	POLN_EEVV3
22	51	19.0	2492	1	POLN_EEVVP
23	50.5	18.8	409	1	AXIA_BRARE
24	50.5	18.8	554	1	DCS2_GOSAR
25	50.5	18.8	2472	1	NCR2_MOUSE
26	50	18.7	130	1	ITP_SCHGR
27	50	18.7	134	1	ITPL_SCHGR
28	50	18.7	456	1	HMFT_DROHY
29	50	18.7	467	1	RBL_PHORE
30	50	18.7	528	1	ANW3_RAT
31	50	18.7	1100	1	JAK3_RAT
32	50	18.7	1163	1	ITAX_HUMAN
33	50	18.7	2492	1	POLN_EEVVT

34	49.5	18.5	490	1	CPCC_RAT
35	49.5	18.5	499	1	RPB2_METVA
36	49.5	18.5	554	1	DCS1_GOSHI
37	49.5	18.5	741	1	IDH2_VIBAL
38	49.5	18.5	893	1	PER_PERAM
39	49	18.3	221	1	SCAB_SHEEP
40	49	18.3	405	1	PE21_RAT
41	49	18.3	608	1	GLMS_ECOLI
42	48.5	18.1	548	1	ILVG_ECOLI
43	48.5	18.1	554	1	DCS1_GOSAR
44	48.5	18.1	644	1	RNR_HELPJ
45	48.5	18.1	834	1	CASL_HUMAN

ALIGNMENTS

RESULT 1	
TA2R_BOVIN	
ID TA2R_BOVIN STANDARD; PRT; 343 AA.	
AC Q95125; Q95124;	
DT 01-NOV-1997 (Rel. 35, Created)	
DT 01-NOV-1997 (Rel. 35, Last sequence update)	
DT 15-JUL-1999 (Rel. 38, Last annotation update)	
DE THROMBOXANE A2 RECEPTOR (TXA2-R) (PROSTANOID TP RECEPTOR).	
GN TBXA2R.	
OS Bos taurus (Bovine).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;	
OC Bovidae; Bovinae; Bos.	
OX NCBI_TaxID=9913;	
RN [1]	
RP SEQUENCE FROM N.A.	
RP TISSUE=Liver, and Heart;	
RC MEDLINE=98221872; PubMed=9561101;	
RX Muck S., Schroer K.;	
RT "Cloning, tissue-specific expression and regulation of the bovine	
RT thromboxane A2 receptor.";	
RL Adv. Exp. Med. Biol. 433:47-50(1997).	
CC -1- FUNCTION: RECEPTOR FOR THROMBOXANE A2 (TXA2), A POTENT STIMULATOR	
CC OF PLATELET AGGREGATION. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED	
CC BY A G-PROTEIN THAT ACTIVATE A PHOSPHATIDYLINOSITOL-CALCIUM SECOND	
CC MESSENGER SYSTEM. IN THE KIDNEY, THE BINDING OF TXA2 TO GLOMERULAR	
CC TP RECEPTORS CAUSES INTENSE VASOCONSTRICTION.	
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.	
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.	
CC This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC the European Bioinformatics Institute. There are no restrictions on its	
CC use by non-profit institutions as long as its content is in no way	
CC modified and this statement is not removed. Usage by and for commercial	
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/	
CC or send an email to license@isb-sib.ch).	
CC -----	
CC EMBL; U53485; AAC34309.1; -	
DR EMBL; U53484; AAC34308.1; -	
DR GCRdb; GCR_1217; -	
DR InterPro; IPR000276; -	
DR InterPro; IPR001105; -	
DR Pfam; PF00001; 7tm_1; 1.	
DR PRINTS; PR00429; THROMBOXANER.	
DR PROSITE; PS00237; G-PROTEIN_RECP_FL_1; 1.	
DR PROSITE; PS0262; G-PROTEIN_RECP_FL_2; 1.	
KW G-protein coupled receptor; Transmembrane; Glycoprotein.	
FT DOMAIN 1 29	
FT TRANSMEM 30 52	
FT DOMAIN 53 66	
FT TRANSMEM 67 87	
FT DOMAIN 88 106	
FT TRANSMEM 107 128	
FT DOMAIN 129 149	
FT TRANSMEM 150 172	
FT DOMAIN 173 194	
FT TRANSMEM 195 216	
FT DOMAIN 217 238	
FT TRANSMEM 239 260	
FT DOMAIN 261 282	
FT TRANSMEM 283 304	
FT DOMAIN 305 326	
FT TRANSMEM 327 348	
FT DOMAIN 349 370	
FT TRANSMEM 371 392	
FT DOMAIN 393 414	
FT TRANSMEM 415 436	
FT DOMAIN 437 458	
FT TRANSMEM 459 480	
FT DOMAIN 481 502	
FT TRANSMEM 503 524	
FT DOMAIN 525 546	
FT TRANSMEM 547 568	
FT DOMAIN 569 590	
FT TRANSMEM 591 612	
FT DOMAIN 613 634	
FT TRANSMEM 635 656	
FT DOMAIN 657 678	
FT TRANSMEM 679 700	
FT DOMAIN 701 722	
FT TRANSMEM 723 744	
FT DOMAIN 745 766	
FT TRANSMEM 767 788	
FT DOMAIN 789 810	
FT TRANSMEM 811 832	
FT DOMAIN 833 854	
FT TRANSMEM 855 876	
FT DOMAIN 877 898	
FT TRANSMEM 899 920	
FT DOMAIN 921 942	
FT TRANSMEM 943 964	
FT DOMAIN 965 986	
FT TRANSMEM 987 1008	
FT DOMAIN 1009 1030	
FT TRANSMEM 1031 1052	
FT DOMAIN 1053 1074	
FT TRANSMEM 1075 1096	
FT DOMAIN 1097 1118	
FT TRANSMEM 1119 1140	
FT DOMAIN 1141 1162	
FT TRANSMEM 1163 1184	
FT DOMAIN 1185 1206	
FT TRANSMEM 1207 1228	
FT DOMAIN 1229 1250	
FT TRANSMEM 1251 1272	
FT DOMAIN 1273 1294	
FT TRANSMEM 1295 1316	
FT DOMAIN 1317 1338	
FT TRANSMEM 1339 1360	
FT DOMAIN 1361 1382	
FT TRANSMEM 1383 1404	
FT DOMAIN 1405 1426	
FT TRANSMEM 1427 1448	
FT DOMAIN 1449 1470	
FT TRANSMEM 1471 1492	
FT DOMAIN 1493 1514	
FT TRANSMEM 1515 1536	
FT DOMAIN 1537 1558	
FT TRANSMEM 1559 1580	
FT DOMAIN 1581 1602	
FT TRANSMEM 1603 1624	
FT DOMAIN 1625 1646	
FT TRANSMEM 1647 1668	
FT DOMAIN 1669 1690	
FT TRANSMEM 1691 1712	
FT DOMAIN 1713 1734	
FT TRANSMEM 1735 1756	
FT DOMAIN 1757 1778	
FT TRANSMEM 1779 1800	
FT DOMAIN 1801 1822	
FT TRANSMEM 1823 1844	
FT DOMAIN 1845 1866	
FT TRANSMEM 1867 1888	
FT DOMAIN 1889 1910	
FT TRANSMEM 1911 1932	
FT DOMAIN 1933 1954	
FT TRANSMEM 1955 1976	
FT DOMAIN 1977 1998	
FT TRANSMEM 1999 2020	
FT DOMAIN 2021 2042	
FT TRANSMEM 2043 2064	
FT DOMAIN 2065 2086	
FT TRANSMEM 2087 2108	
FT DOMAIN 2109 2130	
FT TRANSMEM 2131 2152	
FT DOMAIN 2153 2174	
FT TRANSMEM 2175 2196	
FT DOMAIN 2197 2218	
FT TRANSMEM 2219 2240	
FT DOMAIN 2241 2262	
FT TRANSMEM 2263 2284	
FT DOMAIN 2285 2306	
FT TRANSMEM 2307 2328	
FT DOMAIN 2329 2350	
FT TRANSMEM 2351 2372	
FT DOMAIN 2373 2394	
FT TRANSMEM 2395 2416	
FT DOMAIN 2417 2438	
FT TRANSMEM 2439 2460	
FT DOMAIN 2461 2482	
FT TRANSMEM 2483 2504	
FT DOMAIN 2505 2526	
FT TRANSMEM 2527 2548	
FT DOMAIN 2549 2570	
FT TRANSMEM 2571 2592	
FT DOMAIN 2593 2614	
FT TRANSMEM 2615 2636	
FT DOMAIN 2637 2658	
FT TRANSMEM 2659 2680	
FT DOMAIN 2681 2702	
FT TRANSMEM 2703 2724	
FT DOMAIN 2725 2746	
FT TRANSMEM 2747 2768	
FT DOMAIN 2769 2790	
FT TRANSMEM 2791 2812	
FT DOMAIN 2813 2834	
FT TRANSMEM 2835 2856	
FT DOMAIN 2857 2878	
FT TRANSMEM 2879 2900	
FT DOMAIN 2901 2922	
FT TRANSMEM 2923 2944	
FT DOMAIN 2945 2966	
FT TRANSMEM 2967 2988	
FT DOMAIN 2989 3010	
FT TRANSMEM 3011 3032	
FT DOMAIN 3033 3054	
FT TRANSMEM 3055 3076	
FT DOMAIN 3077 3098	
FT TRANSMEM 3099 3120	
FT DOMAIN 3121 3142	
FT TRANSMEM 3143 3164	
FT DOMAIN 3165 3186	
FT TRANSMEM 3187 3208	
FT DOMAIN 3209 3230	
FT TRANSMEM 3231 3252	
FT DOMAIN 3253 3274	
FT TRANSMEM 3275 3296	
FT DOMAIN 3297 3318	
FT TRANSMEM 3319 3340	
FT DOMAIN 3341 3362	
FT TRANSMEM 3363 3384	
FT DOMAIN 3385 3406	
FT TRANSMEM 3407 3428	
FT DOMAIN 3429 3450	
FT TRANSMEM 3451 3472	
FT DOMAIN 3473 3494	
FT TRANSMEM 3495 3516	
FT DOMAIN 3517 3538	
FT TRANSMEM 3539 3560	
FT DOMAIN 3561 3582	
FT TRANSMEM 3583 3604	
FT DOMAIN 3605 3626	
FT TRANSMEM 3627 3648	
FT DOMAIN 3649 3670	
FT TRANSMEM 3671 3692	
FT DOMAIN 3693 3714	
FT TRANSMEM 3715 3736	
FT DOMAIN 3737 3758	
FT TRANSMEM 3759 3780	
FT DOMAIN 3781 3802	
FT TRANSMEM 3803 3824	
FT DOMAIN 3825 3846	
FT TRANSMEM 3847 3868	
FT DOMAIN 3869 3890	
FT TRANSMEM 3891 3912	
FT DOMAIN 3913 3934	
FT TRANSMEM 3935 3956	
FT DOMAIN 3957 3978	
FT TRANSMEM 3979 4000	
FT DOMAIN 4001 4022	
FT TRANSMEM 4023 4044	
FT DOMAIN 4045 4066	
FT TRANSMEM 4067 4088	
FT DOMAIN 4089 4110	
FT TRANSMEM 4111 4132	
FT DOMAIN 4133 4154	
FT TRANSMEM 4155 4176	
FT DOMAIN 4177 4198	
FT TRANSMEM 4199 4220	
FT DOMAIN 4221 4242	
FT TRANSMEM 4243 4264	
FT DOMAIN 4265 4286	
FT TRANSMEM 4287 4308	
FT DOMAIN 4309 4330	
FT TRANSMEM 4331 4352	
FT DOMAIN 4353 4374	
FT TRANSMEM 4375 4396	
FT DOMAIN 4397 4418	
FT TRANSMEM 4419 4440	
FT DOMAIN 4441 4462	
FT TRANSMEM 4463 4484	
FT DOMAIN 4485 4506	
FT TRANSMEM 4507 4528	
FT DOMAIN 4529 4550	
FT TRANSMEM 4551 4572	
FT DOMAIN 4573 4594	
FT TRANSMEM 4595 4616	
FT DOMAIN 4617 4638	
FT TRANSMEM 4639 4660	
FT DOMAIN 4661 4682	
FT TRANSMEM 4683 4704	
FT DOMAIN 4705 4726	
FT TRANSMEM 4727 4748	
FT DOMAIN 4749 4770	
FT TRANSMEM 4771 4792	
FT DOMAIN 4793 4814	
FT TRANSMEM 4815 4836	
FT DOMAIN 4837 4858	
FT TRANSMEM 4859 4880	
FT DOMAIN 4881 4902	
FT TRANSMEM 4903 4924	
FT DOMAIN 4925 4946	
FT TRANSMEM 4947 4968	
FT DOMAIN 4969 4990	
FT TRANSMEM 4991 5012	
FT DOMAIN 5013 5034	
FT TRANSMEM 5035 5056	
FT DOMAIN 5057 5078	
FT TRANSMEM 5079 5100	
FT DOMAIN 5101 5122	
FT TRANSMEM 5123 5144	
FT DOMAIN 5145 5166	
FT TRANSMEM 5167 5188	
FT DOMAIN 5189 5210	
FT TRANSMEM 5211 5232	
FT DOMAIN 5233 5254	
FT TRANSMEM 5255 5276	
FT DOMAIN 5277 5298	
FT TRANSMEM 5299 5320	
FT DOMAIN 5321 5342	
FT TRANSMEM 5343 5364	
FT DOMAIN 5365 5386	
FT TRANSMEM 5387 5408	
FT DOMAIN 5409 5430	
FT TRANSMEM 5431 5452	
FT DOMAIN 5453 5474	
FT TRANSMEM 5475 5496	
FT DOMAIN 5497 5518	
FT TRANSMEM 5519 5540	
FT DOMAIN 5541 5562	
FT TRANSMEM 5563 5584	
FT DOMAIN 5585 5606	
FT TRANSMEM 5607 5628	
FT DOMAIN 5629 5650	
FT TRANSMEM 5651 5672	
FT DOMAIN 5673 5694	
FT TRANSMEM 5695 5716	
FT DOMAIN 5717 5738	
FT TRANSMEM 5739 5760	
FT DOMAIN 5761 5782	
FT TRANSMEM 5783 5804	
FT DOMAIN 5805 5826	
FT TRANSMEM 5827 5848	
FT DOMAIN 5849 5870	
FT TRANSMEM 5871 5892	
FT DOMAIN 5893 5914	
FT TRANSMEM 5915 5936	
FT DOMAIN 5937 5958	
FT TRANSMEM 5959 5980	
FT DOMAIN 5981 6002	
FT TRANSMEM 6003 6024	
FT DOMAIN 6025 6046	
FT TRANSMEM 6047 6068	
FT DOMAIN 6069 6090	
FT TRANSMEM 6091 6112	
FT DOMAIN 6113 6134	
FT TRANSMEM 6135 6156	
FT DOMAIN 6157 6178	
FT TRANSMEM 6179 6200	
FT DOMAIN 6201 6222	
FT TRANSMEM 6223 6244	
FT DOMAIN 6245 6266	
FT TRANSMEM 6267 6288	
FT DOMAIN 6289 6310	
FT TRANSMEM 6311 6332	
FT DOMAIN 6333 6354	
FT TRANSMEM 6355 6376	
FT DOMAIN 6377 6398	
FT TRANSMEM 6399 6420	
FT DOMAIN 6421 6442	
FT TRANSMEM 6443 6464	
FT DOMAIN 6465 6486	
FT TRANSMEM 6487 6508	
FT DOMAIN 6509 6530	
FT TRANSMEM 6531 6552	
FT DOMAIN 6553 6574	
FT TRANSMEM 6575 6596	
FT DOMAIN 6597 6618	
FT TRANSMEM 6619 6640	
FT DOMAIN 6641 6662	
FT TRANSMEM 6663 6684	
FT DOMAIN 6685 6706	
FT TRANSMEM 6707 6728	
FT DOMAIN 6729 6750	
FT TRANSMEM 6751 6772	
FT DOMAIN 6773 6794	
FT TRANSMEM 6795 6816	
FT DOMAIN 6817 6838	
FT TRANSMEM 6839 6860	
FT DOMAIN 6861 6882	
FT TRANSMEM 6883 6904	
FT DOMAIN 6905 6926	
FT TRANSMEM 6927 6948	
FT DOMAIN 6949 6970	
FT TRANSMEM 6971 6992	
FT DOMAIN 6993 7014	
FT TRANSMEM 7015 7036	
FT DOMAIN 7037 7058	
FT TRANSMEM 7059 7080	
FT DOMAIN 7081 7102	
FT TRANSMEM 7103 7124	
FT DOMAIN 7125 7146	
FT TRANSMEM 7147 7168	
FT DOMAIN 7169 7190	
FT TRANSMEM 7191 7212	
FT DOMAIN 7213 7234	
FT TRANSMEM 7235 7256	
FT DOMAIN 7257 7278	
FT TRANSMEM 7279 7300	
FT DOMAIN 7301 7322	
FT TRANSMEM 7323 7344	
FT DOMAIN 7345 7366	
FT TRANSMEM 7367 7388	
FT DOMAIN 7389 7410	
FT TRANSMEM 7411 7432	
FT DOMAIN 7433 7454	
FT TRANSMEM 7455 7476	
FT DOMAIN 7477 7498	
FT TRANSMEM 7499 7520	
FT DOMAIN 7521 7542	
FT TRANSMEM 7543 7564	
FT DOMAIN 7565 7586	
FT TRANSMEM 7587 7608	
FT DOMAIN 7609 7630	
FT TRANSMEM 7631 7652	
FT DOMAIN 7653 7674	
FT TRANSMEM 7675 7696	
FT DOMAIN 7697 7718	
FT TRANSMEM 7719 7740	
FT DOMAIN 7741 7762	
FT TRANSMEM 7763 7784	
FT DOMAIN 7785 7806	
FT TRANSMEM 7807 7828	
FT DOMAIN 7829 7850	
FT TRANSMEM 7851 7872	
FT DOMAIN 7873 7894	
FT TRANSMEM 7895 7916	
FT DOMAIN 7917 7938	
FT TRANSMEM 7939 7960	
FT DOMAIN 7961 7982	
FT TRANSMEM 7983 8004	
FT DOMAIN 8005 8026	
FT TRANSMEM 8027 8048	
FT DOMAIN 8049 8070	
FT TRANSMEM 8071 8092	
FT DOMAIN 8093 8114	
FT TRANSMEM 8115 8136	
FT DOMAIN 8137 8158	
FT TRANSMEM 8159 8180	
FT DOMAIN 8181 8202	
FT TRANSMEM 8203 8224	
FT DOMAIN 8225 8246	
FT TRANSMEM 8247 8268	
FT DOMAIN 8269 8290	
FT TRANSMEM 8291 8312	
FT DOMAIN 8313 8334	
FT TRANSMEM 8335 8356	
FT DOMAIN 8357 8378	
FT TRANSMEM 8379 8400	
FT DOMAIN 8401 8422	
FT TRANSMEM 8423 8444	
FT DOMAIN 8445 8466	
FT TRANSMEM 8467 8488	
FT DOMAIN 8489 8510	
FT TRANSMEM 8511 8532	
FT DOMAIN 8533 8554	
FT TRANSMEM 8555 8576	
FT DOMAIN 8577 8598	
FT TRANSMEM 8599 8620	
FT DOMAIN 8621 8642	
FT TRANSMEM 8643 8664	
FT DOMAIN 8665 8686	
FT TRANSMEM 8687 8708	
FT DOMAIN 8709 8730	
FT TRANSMEM 8731 8752	
FT DOMAIN 8753 8774	
FT TRANSMEM 8775 8796	
FT DOMAIN 8797 8818	
FT TRANSMEM 8819 8840	
FT DOMAIN 8841 8862	
FT TRANSMEM 8863 8884	
FT DOMAIN 8885 8906	
FT TRANSMEM 8907 8928	
FT DOMAIN 8929 8950	
FT TRANSMEM 8951 8972	
FT DOMAIN 8973 8994	
FT TRANSMEM 8995 9016	
FT DOMAIN 9017 9038	
FT TRANSMEM 9039 9060	
FT DOMAIN 9061 9082	
FT TRANSMEM 9083 9104	
FT DOMAIN 9105 9126	
FT TRANSMEM 9127 9148	
FT DOMAIN 9149 9170	
FT TRANSMEM 9171 9192	
FT DOMAIN 9193 9214	
FT TRANSMEM 9215 9236	
FT DOMAIN 9237 9258	
FT TRANSMEM 9259 9280	
FT DOMAIN 9281 9302	
FT TRANSMEM 9303 9324	
FT DOMAIN 9325 9346	
FT TRANSMEM 9347 9368	
FT DOMAIN	

Hydrolase; Nuclear protein.	POLY-GLY.	POLY-GLN.	POLY-GLU.	HISTONE DEACETYLASE.
DOMAIN 47	52			
FT				
DOMAIN 85	92			
FT				
DOMAIN 577	588			
FT				
DOMAIN 675	1019			
FT				
SEQUENCE 1113	AA; 120942	MW; 63071	AA; 120942	MW; 63071
SEQ				

DR PRINTS; PR01270; HDASUPER.

DR PRINTS; PR01270; HDASUPER.

GN CORO1A OR CORO1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Spleen;
RX MEDLINE=95278344; PubMed=7758584;
RA Suzuki K., Nishihata J., Arai Y., Honma N., Yamamoto K., Irimura T.,
Toyoshima S.;
RT "Molecular cloning of a novel actin-binding protein, p57, with a WD
repeat and a leucine zipper motif.";
RL FEBS Lett. 364:283-288(1995).
CC -!- FUNCTION: MAY PLAY A ROLE IN THE SIGNAL TRANSDUCTION PATHWAYS OF
CHEMOTAXIS.
CC -!- SUBUNIT: BINDS ACTIN.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, THYMUS, SPLEEN, BONE
MARROW AND LYMPH NODE. LOW IN LUNG AND GUT.
CC -!- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
CC -!- SIMILARITY: BELONGS TO THE CORONIN FAMILY OF WD-REPEAT PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; D44496; BAA07939.1; -
CC InterPro: IPR001680; -
CC Pfam: PF00400; WD40; 3.
CC PROSITE; PS00678; WD_REPEATS_1; 2.
CC PROSITE; PS50082; WD_REPEATS_2; 2.
CC PROSITE; PS50294; WD_REPEATS_REGION; 1.
CC Actin-binding; Repeat; WD repeat; Coiled coil.
KW REPEAT 79 119 WD 1.
FT REPEAT 129 169 WD 2.
FT REPEAT 174 213 WD 3.
FT REPEAT 218 260 WD 4.
FT REPEAT 265 305 WD 5.
FT DOMAIN 424 460 COILED COIL (POTENTIAL).
FT SEQUENCE 461 AA; 50979 MW; 1721A5D093C1130A CRC64;
SQ
Query Match 20.3%; Score 54.5; DB 1; Length 461;
Best Local Similarity 43.8%; Pred. No. 13;
Matches 14; Conservative 3; Mismatches 8; Indels 7; Gaps 1;
OY 2 FLHISPPKYPHTQEAQKQASRLGEMPGRLH 33
DB 303 FLHLSMF-----SSKESQRMGYMPKRG 327
RESULT 6
COLA.HUMAN
ID COLA.HUMAN STANDARD; PRT; 461 AA.
AC P31146;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CORONIN-LIKE PROTEIN P57 (CORONIN 1A).
GN CORO1A OR CORO1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Peripheral blood leukocytes;
RX MEDLINE=95278344; PubMed=7758584;

RA Suzuki K., Nishihata J., Arai Y., Honma N., Yamamoto K., Irimura T.,
Toyoshima S.;
RT "Molecular cloning of a novel actin-binding protein, p57, with a WD
repeat and a leucine zipper motif.";
RL FEBS Lett. 364:283-288(1995).
CC -!- FUNCTION: MAY PLAY A ROLE IN THE SIGNAL TRANSDUCTION PATHWAYS OF
CHEMOTAXIS.
CC -!- SUBUNIT: BINDS ACTIN (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, THYMUS, SPLEEN, BONE
MARROW AND LYMPH NODE. LOW IN LUNG AND GUT.
CC -!- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
CC -!- SIMILARITY: BELONGS TO THE CORONIN FAMILY OF WD-REPEAT PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; D44497; BAA07940.1; -
CC EMBL; X89109; CAA61482.1; -
CC EMBL; U34690; AAA77058.1; -
CC MIM; 605000; -
CC Aarhus/Ghent-2DPAGE; 2416; IEF.
CC InterPro: IPR001680; -
CC Pfam: PF00400; WD40; 3.
CC PROSITE; PS00678; WD_REPEATS_1; 2.
CC PROSITE; PS50082; WD_REPEATS_2; 2.
CC PROSITE; PS50294; WD_REPEATS_REGION; 1.
CC Actin-binding; Repeat; WD repeat; Coiled coil.
KW REPEAT 79 119 WD 1.
FT REPEAT 129 169 WD 2.
FT REPEAT 174 213 WD 3.
FT REPEAT 218 260 WD 4.
FT REPEAT 265 305 WD 5.
FT DOMAIN 424 460 COILED COIL (POTENTIAL).
FT VARIANT 355 355 K->D.
FT /FTID=VAR_007722.
FT S->D.
FT VARIANT 356 356
FT /FTID=VAR_007723.
FT CONFLICT 8 8 S->T (IN REF. 3).
FT CONFLICT 245 245 R->W (IN REF. 3).
SQ SEQUENCE 461 AA; 51026 MW; DE3FEDA57041515E CRC64;
Query Match 20.3%; Score 54.5; DB 1; Length 461;
Best Local Similarity 43.8%; Pred. No. 13;
Matches 14; Conservative 3; Mismatches 8; Indels 7; Gaps 1;
OY 2 FLHISPPKYPHTQEAQKQASRLGEMPGRLH 33
DB 303 FLHLSMF-----SSKESQRMGYMPKRG 327

```

RESULT 7
COIA_MOUSE
ID COIA_MOUSE STANDARD; PRT; 461 AA.
AC O89053; Q9R288;
DT 30-MAY-2000 (Rel. 39, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CORONIN-LIKE PROTEIN P57 (CORONIN 1A).
GN COROLA OR COROL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98449467; PubMed=9778037;
RA Okumura M., Kung C., Wong S., Rodgers M., Thomas M.L.;
RT "Definition of family of coronin-related proteins conserved between
humans and mice: close genetic linkage between coronin-2 and CD45-
associated protein."
RT associated protein."
RL DNA Cell Biol. 17:779-787(1998).
RN [2]
RP SEQUENCE OF 156-276 FROM N.A.
RC STRAIN=BALB/C; TISSUE=Spleen;
RX MEDLINE=99012977; PubMed=9798653;
RA Chu C.C., Paul W.E.;
RT "Expressed genes in interleukin-4 treated B cells identified by cDNA
representational difference analysis."
RL Mol. Immunol. 35:487-502(1998).
CC -!- FUNCTION: MAY PLAY A ROLE IN THE SIGNAL TRANSDUCTION PATHWAYS OF
CHEMOTAXIS (BY SIMILARITY).
CC -!- SUBUNIT: BINDS ACTIN (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: EXPRESSION RESTRICTED TO HEMATOPOIETIC CELLS.
CC -!- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
CC -!- SIMILARITY: BELONGS TO THE CORONIN FAMILY OF WD-REPEAT PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
DR EMBL; AF143955; AAD32703.1;
DR EMBL; U89399; AAC36506.1;
DR MGD; MGI:1345961; Corola.
DR InterPro; IPR001680;
DR Pfam; PF00400; WD40; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 2.
DR PROSITE; PS00082; WD_REPEATS_2; 2.
DR PROSITE; PS02994; WD_REPEATS_REGION; 1.
KW Actin-binding; Repeat; WD repeat; Coiled coil.
FT REPEAT 79 119 WD 1.
FT REPEAT 129 169 WD 2.
FT REPEAT 174 213 WD 3.
FT REPEAT 218 260 WD 4.
FT REPEAT 265 305 WD 5.
FT DOMAIN 424 460 COILED COIL (POTENTIAL).
FT SEQUENCE 461 AA; 50975 MW; 51D48ED91E6FE82A CRC64;

Query Match 20.3%; Score 54.5; DB 1; Length 461;
Best Local Similarity 43.8%; Pred. NO. 13;
Matches 14; Conservative 3; Mismatches 8; Indels 7; Gaps 1;

QY 2 FLHISPFKYPHTQEAQKQSLGEMPGRHL 33
| | | | : | | | | | | | |
DB 303 FLHLSMF-----SSKRSQRMGMYPKRG 327

```

```

RESULT 8
GNTU_ECOLI
ID GNTU_ECOLI STANDARD; PRT; 445 AA.
AC P46858; P76694; P76695;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE LOW-AFFINITY GLUCONATE TRANSPORTER (GLUCONATE PERMEASE) (GNT-I
SYSTEM).
GN GNTU OR B3435/B3436.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=97280784; PubMed=9135111;
RA Izu H., Adachi O., Yamada M.;
RT "Gene organization and transcriptional regulation of the gntRku
operon involved in gluconate uptake and catabolism of Escherichia
coli."
RT coli."
RL J. Mol. Biol. 267:778-793(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=96235044; PubMed=8655507;
RA Tong S., Porco A., Istriz T., Conway T.;
RT "Cloning and molecular genetic characterization of the Escherichia
coli gntR, gntK, and gntU genes of GntI, the main system for
gluconate metabolism."
RL J. Bacteriol. 178:3260-3269(1996).
CC -!- FUNCTION: PART OF THE GLUCONATE UTILIZATION SYSTEM GNT-I; LOW-
AFFINITY INTAKE OF GLUCONATE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
(POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE GNTP FAMILY OF PERMEASES.
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
FRAMESHIFT.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
DR EMBL; D84362; BAAL2326.1;
DR EMBL; U18997; AAA58234.1; ALT_FRAME.
DR EMBL; U18997; AAA58233.1; ALT_FRAME.
DR EMBL; AE000420; AAC76461.1; ALT_FRAME.
DR EMBL; AE000420; AAC76460.1; ALT_FRAME.
DR EcoGene; EG12631; gntU.
KW Gluconate utilization; Sugar transport; Transmembrane;
KW Inner membrane; Multigene family.
FT TRANSMEM 2 22 POTENTIAL.
FT TRANSMEM 27 47 POTENTIAL.
FT TRANSMEM 59 79 POTENTIAL.
FT TRANSMEM 110 130 POTENTIAL.
FT TRANSMEM 143 163 POTENTIAL.
FT TRANSMEM 177 197 POTENTIAL.
FT TRANSMEM 226 246 POTENTIAL.
FT TRANSMEM 262 282 POTENTIAL.

```

```

Query Match          20.08; Score 53.5; DB 1; Length 451;
Best Local Similarity 28.09; Pred. No. 17;
Matches 14; Conservative 12; Mismatches 19; Indels 5; Gaps 1;

QY      5 ISSPKYPHTOEAKAORSIGEMPRHLGS-----SMSLALCLPLVIRE 49
       ::::| | | | | | | | | | | | | | | | | | | | :
Db      116 LATPYQFPFTPEGAKETIMVGGIPVQWVGSKGLFVAMILAIVSTEIYRK 165
       ::::| | | | | | | | | | | | | | | | | | | | :

RESULT 10
SNTD_BOOMT
ID      SNTD_BOOMI        STANDARD;          PRT;    580 AA.
AC      P52307; P90696;
DT      01-OCT-1996 (Rel. 34, Created)
DD      DT 01-OCT-2000 (Rel. 40, Last sequence update)
DE      01-OCT-2000 (Rel. 40, Last annotation update)
DS      5'-NUCLEOTIDASE PRECURSOR (EC 3.1.3.5) (FRAGMENT).
OS      Boophilus microplus (cattle tick).
OC      Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC      Parasitiformes; Ixodida; Ixodidae; Boophilus.
NCBI_TaxID=6941;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE-99306777; Pubmed=10390109;
RA      Llyou N., Hamilton S., Elvin C., Willadsen P.;
RT      "Cloning and expression of ecto 5-nucleotidase from the cattle tick
RT      Boophilus microplus.";
RL      Insect Mol. Biol. 8:257-266(1999).
[2]
RP      SEQUENCE OF 15-40 AND 162-180.
RX      MEDLINE-93250870; Pubmed=8387372;
RA      Willadsen P., Riding G.A., Jarney J., Atkins A.;
RT      "The nucleotidase of Boophilus microplus and its relationship to
RT      enzymes from the rat and Escherichia coli.";
RL      Insect Biochem. Mol. Biol. 23:291-295(1993).
CC      -|- FUNCTION. DEGRADATION OF EXTERNAL UDP-GLUCOSE TO URIDINE
CC      MONOPHOSPHATE AND GLUCOSE-1-PHOSPHATE, WHICH CAN THEN BE USED BY
CC      THE CELL.
CC      -|- CATALYTIC ACTIVITY: UDP-SUGAR + H(2)O = UMP + SUGAR 1-PHOSPHATE.
CC      -|- CATALYTIC ACTIVITY: A 5'-RIBONUCLEOTIDE + H(2)O = A
CC      RIBONUCLEOSIDE + ORTHOPHOSPHATE.
CC      -|- COFACTOR: ZINC.
CC      -|- SUBUNIT: HOMODIMER.
CC      -|- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC      -|- TISSUE SPECIFICITY: GUT, OVARIES AND SALIVARY GLANDS.
CC      -|- PTM: GLYCOSYLATED.
CC      -|- SIMILARITY: BELONGS TO THE 5'-NUCLEOTIDASE FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation at
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U80634; AAB38963.1; -.
InterPro; IPR000934; -.
InterPro; IPR022224; -.
DR      PROSITE; PS00785; 5_NUCLEOTIDASE_1; FALSE_NEG.
DR      PROSITE; PS00786; 5_NUCLEOTIDASE_2; FALSE_NEG.
DR      Pfam; PF01009; 5_nucleotidase; 1.
KW      Hydrolase; GPI-anchor; Glycoprotein; Signal; Zinc.
FT      NON_TER      1      1
FT      CHAIN         <1      14
FT      PEPTIDE              14
FT      PROPEP           553      580
FT      LIPID             552      552
FT      CARBOHYD          172      172
FT      CARBOHYD          285      285
FT      CARBOHYD          423      423
FT      CARBOHYD          536      536
FT      SIGNAL          15      552
FT      REMOVED IN MATURE FORM (BY SIMILARITY).
FT      GPI-ANCHOR (POTENTIAL).
FT      N-LINKED (GLCNAC. . .) (PROBABLE).
FT      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```
FT CONFLICT 15 15 T -> K (IN REF. 2).
FT CONFLICT 37 39 SGT -> HXG (IN REF. 2).
SQ SEQUENCE 580 AA; 63460 MW; 588EEF2014071AB7 CRC64;

Query Match
Best Local Similarity 36.1%; Score 53.5; DB 1; Length 580;
Matches 22; Conservative 5; Mismatches 21; Indels 13; Gaps 3;

Qy 4 HISSPFKYPHTQEA--OKEAQRSLGE-----MPGRHGLSSMSLALC-----LVPVLYREG 50
Db 175 YLSSPGKVRFTDEACIOREAQRLLREGVQVIVAGVSGVPRDLCEICRVPVSLVVG 234
Qy 51 H 51
Db 235 H 235

RESULT 11
DCS4_GOSAR STANDARD; PRT; 554 AA.
AC 049853; 1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE (+)-DELTA-CADINENE SYNTHASE ISOZYME C2 (EC 4.6.1.11) (D-CADINENE
SYNTHASE).
GN CAD1-C2.
OS Gossypium arboreum.
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Malvales; Malvaceae; Gossypium.
OX NCBI_TaxID=29729;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NANKING;
RA Meng Y., Jia J., Liu C., Liang W., Zhou X., Heinstein P., Chen X.-Y.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: (E,E)-FARNESYL DIPOSPHATE CYCLIZING.
CC -1- CATALYTIC ACTIVITY: 2-TRANS-6-TRANS-FARNESYL DIPOSPHATE = (+)-
DELTA-CADINENE + DIPHOSPHATE.
CC -1- PATHWAY: FIRST COMMITTED STEP IN THE PATHWAYS LEADING TO THE
RELATED PHYTOALEXINS GOSSYPOL AND LACINILENE C.
CC -1- SIMILARITY: TO OTHER PLANT TERPENE CYCLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y16432; CAA76223.1; -.
DR HSSP; Q40577; 5EAU.
DR InterPro; IPR001906; -.
DR Pfam; PF01397; Terpene_synth; 1.
KW Lyase; Multigene family.
SQ SEQUENCE 554 AA; 64117 MW; 35DD66D6D3E838AAC CRC64;

Query Match
Best Local Similarity 38.3%; Score 53; DB 1; Length 554;
Matches 18; Conservative 5; Mismatches 18; Indels 6; Gaps 2;

Qy 4 HISSPFKYPHTQEAQRSLGEMPGRHGLSSMSLALCLVPLVREG 50
Db 490 HVESAWK-DVNOEQKPT-----EMPTVLRSLNLRVMDVLYREG 530

RESULT 12
RXRA_BRARE STANDARD; PRT; 441 AA.
ID RXRA_BRARE
```

```
Q90416;
AC 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE RETINOIC ACID RECEPTOR RXR-ALPHA.
GN RXRA OR NR2B1 OR RXR.
OS Brachydanio rerio (zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96009547; PubMed=7565671;
RA Jones B.B., Ohno C.K., Allenby G., Boffa M.B., Levin A.A.,
RA Grippo J.F., Petkovich M.;
RT "New retinoid X receptor subtypes in zebra fish (Danio rerio)
RT differentially modulate transcription and do not bind 9-cis retinoic
acid.";
RL Mol. Cell. Biol. 15:5226-5234(1995).
CC -1- FUNCTION: INVOLVED IN RETINOIC ACID RESPONSE PATHWAY. BINDS
9-CIS RETINOIC ACID (9C-RA) (BY SIMILARITY).
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC NR2 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U29940; AAC59720.1; -.
DR HSSP; P19793; 2NLL.
DR ZFIN; ZDB-GENE-980526-36; rxra.
DR InterPro; IPR000003; -.
DR InterPro; IPR000536; -.
DR InterPro; IPR001628; -.
DR InterPro; IPR001723; -.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00047; STROIDFINGER.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00545; RETINOIDXR.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
Zinc-finger; Multigene family.
KW DOMAIN 1 116 MODULATING (BY SIMILARITY).
FT DNA_BIND 117 182 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 117 137 C4-TYPE.
FT ZN_FING 153 177 C4-TYPE.
FT DOMAIN 183 206 HINGE.
FT DOMAIN 207 441 LIGAND-BINDING (BY SIMILARITY).
SQ SEQUENCE 441 AA; 48690 MW; E0F8F1EBE31CA1EC CRC64;

Query Match
Best Local Similarity 31.1%; Score 52.5; DB 1; Length 441;
Matches 14; Conservative 6; Mismatches 18; Indels 7; Gaps 1;

Qy 2 FLHISSPFKYPHTQEAQRSLGEMPGRH-----LGSSMSL 39
Db 7 YLHLSSSLQVAHGLSSPPSPPLGLSMVSHHHPHSLGLSPSY 51

RESULT 13
RXR2_TRYBB STANDARD; PRT; 337 AA.
ID RXR2_TRYBB
```

AC 015910; 015880;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE SMALL CHAIN (EC 1.17.4.1)
DE (RIBONUCLEOTIDE REDUCTASE R2 SUBUNIT).
GN RN2 OR NRDB.
OS Trypanosoma brucei brucei.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5702;
RN (1)
RN SEQUENCE FROM N.A.
RP STRAIN=427;
RX MEDLINE=97338127; PubMed=9192674;
RA Hofer A., Schmidt P.P., Graslund A., Thelander L.;
RT "Cloning and characterization of the R1 and R2 subunits of
RT ribonucleotide reductase from Trypanosoma brucei";
RL Proc. Natl. Acad. Sci. U.S.A. 94:6959-6964(1997).
RN [2]
RN SEQUENCE FROM N.A.
RP MEDLINE=97459778; PubMed=9315738;
RX Dornmeyer M., Schoneck R., Dittmar G.A.G., Krauth-Siegel R.L.;
RA "Cloning, sequencing and expression of ribonucleotide reductase R2
RT from Trypanosoma brucei";
RT FEBS Lett. 414:449-453(1997).
CC -!- FUNCTION: PROVIDES THE PRECURSORS NECESSARY FOR DNA SYNTHESIS.
CC -!- CATALYTIC ACTIVITY: 2'DEoxyRIBONUCLEOSIDE DIPHOSPHATE + OXIDIZED
CC THIOREDOXIN + H(2)O = RIBONUCLEOSIDE DIPHOSPHATE + REDUCED
CC THIOREDOXIN.
CC -!- COFACTOR: CONTAINS TWO IRON IONS.
CC -!- PATHWAY: FIRST REDIMER IN THE DNA REPLICATION PATHWAY.
CC -!- SUBUNIT: HETERODIMER OF A LARGE AND A SMALL CHAIN.
CC -!- SIMILARITY: BELONGS TO THE RIBONUCLEOSIDE DIPHOSPHATE REDUCTASE
CC SMALL CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U80911; AAB70705.1; -
DR EMBL; Y10768; CAA71741.1; -
DR InterPro; IPR000358; -
DR Pfam; PF00268; ribonuc_red.1.
DR PROSITE; PS00368; RIBRED_SMALL; 1.
KW Oxidoreductase; DNA replication; Iron.
FT METAL 85 85 IRON 1 (BY SIMILARITY).
FT METAL 116 116 IRON 1 AND 2 (BY SIMILARITY).
FT METAL 119 119 IRON 1 (BY SIMILARITY).
FT METAL 178 178 IRON 2 (BY SIMILARITY).
FT METAL 212 212 IRON 2 (BY SIMILARITY).
FT METAL 215 215 IRON 2 (BY SIMILARITY).
FT ACT_SITE 123 123 BY SIMILARITY.
FT CONFLICT 8 10 CSR -> RSA (IN REF. 2).
FT CONFLICT 43 43 E -> K (IN REF. 2).
SQ SEQUENCE 337 AA; 39008 MW; BD39919280C65CAE CRC64;

Query Match 19.4%; Score 52; DB 1; Length 337;
Best Local Similarity 31.4%; Pred. No. 20;
Matches 11; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

QY 4 HISSPPKYPHTQEAQKRSLSGEMPGRHLSMS 38
::: ||||: |||: | : ||: | :
Db 27 YVIFPKYPIQWKYKEASSIWTVEIDLGNMT 61

RESULT 14
ID MAPA_RAT STANDARD; PRT: 2774 AA.

AC P34926;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE MICROTUBULE-ASSOCIATED PROTEIN 1A [CONTAINS: MAP1 LIGHT CHAIN LC2].
GN MAP1A.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Brain;
RX MEDLINE=92355629; PubMed=1379599;
RA Landkopf A., Hammarback J.A., Mueller R., Vallee R.B., Garner C.C.;
RT "Microtubule-associated proteins 1A and LC2. Two proteins encoded in
RT one messenger RNA";
RL J. Biol. Chem. 267:16561-16566(1992).
CC -!- FUNCTION: STRUCTURAL PROTEIN INVOLVED IN THE FILAMENTOUS
CC CROSS-BRIDGING BETWEEN MICROTUBULES AND OTHER SKELETAL ELEMENTS.
CC -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
CC WITH MAP1A AND MAP1B PROTEINS.
CC -!- TISSUE SPECIFICITY: BRAIN, HEART AND MUSCLE.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED LATE DURING NEURONAL DEVELOPMENT
CC APPEARING WHEN AXONS AND DENDRITES BEGIN TO SOLIDIFY AND STABILIZE
CC THEIR MORPHOLOGY.
CC -!- DOMAIN: THE BASIC REGION CONTAINING THE REPEATS MAY BE RESPONSIBLE
CC FOR THE BINDING OF MAP1A TO MICROTUBULES.
CC -!- PTM: VARIOUS SERINE RESIDUES MAY BE PHOSPHORYLATED BY CAMP KINASE.
CC -!- PTM: LC2 IS COEXPRESSED WITH MAP1A. IT IS A POLYPEPTIDE GENERATED
CC FROM MAP1A BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
CC BOTH MAP1A AND MAP1B.
CC -!- SIMILARITY: TO MAP1B.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M83196; AAB48069.1; -
DR PIR; A43359; A43359.
KW Microtubules; Repeat; Phosphorylation.
FT CHAIN 72465 2774 MAP1 LIGHT CHAIN LC2.
FT DOMAIN 309 496 LYS-RICH (BASIC).
FT DOMAIN 336 541 11 X 3 AA REPEATS OF K-K-[DE].
FT REPEAT 336 338 1.
FT REPEAT 415 417 2.
FT REPEAT 420 422 3.
FT REPEAT 424 426 4.
FT REPEAT 427 429 5.
FT REPEAT 431 433 6.
FT REPEAT 436 438 7.
FT REPEAT 440 442 8.
FT REPEAT 444 446 9.
FT REPEAT 449 451 10.
FT REPEAT 539 541 11.
SQ SEQUENCE 2774 AA; 299526 MW; 3DEF74427BA9D7D7 CRC64;

Query Match 19.4%; Score 52; DB 1; Length 2774;
Best Local Similarity 46.2%; Pred. No. 1.8e+02;
Matches 12; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 9 FKYPHTQEAQKRSLSGEMPGRHLSG 34
| ||||| : | : ||||
Db 678 FYQKHTQEAQKRSLSGEMPGRHLSG 703

RESULT 15
AQPI_HUMAN

AD AQP1_HUMAN STANDARD; PRT; 269 AA.
AC P29972;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE AQPAPORIN-CHIP (WATER CHANNEL PROTEIN FOR RED BLOOD CELLS AND KIDNEY
DE PROXIMAL TUBULE) (AQPAPORIN 1) (AQP-1) (URINE WATER CHANNEL).
GN AQP1 OR CHIP28.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=92107900; PubMed=1722319;
RA Preston G.M., Agre P.;
RT "Isolation of the cDNA for erythrocyte integral membrane protein of
RT 28 kilodaltons: member of an ancient channel family.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:11110-11114(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93340184; PubMed=8340403;
RA Moon C., Preston G.M., Griffin C.A., Jabs E.W., Agre P.;
RT "The human aquaporin-CHIP gene. Structure, organization, and
RT chromosomal localization.";
RL J. Biol. Chem. 268:15772-15778(1993).
RN [3]
RP SEQUENCE FROM N.A.
RX TISSUE=Retinal pigment epithelium;
RC MEDLINE=96326579; PubMed=8703970;
RA Ruiz A.C., Bok D.;
RT "Characterization of the 3' UTR sequence encoded by the AQP-1 gene in
RT human retinal pigment epithelium.";
RL Biochim. Biophys. Acta 1282:174-178(1996).
RN [4]
RP SEQUENCE FROM N.A.
RX TISSUE=Uterus;
RC MEDLINE=94290349; PubMed=7517253;
RA Li X., Yu H., Koide S.S.;
RT "The water channel gene in human uterus.";
RL Biochem. Mol. Biol. Int. 32:371-377(1994).
RN [5]
RP SEQUENCE OF 1-128 FROM N.A.
RA Dempsey S., Lacy M., Holmes A., Nguyen C.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 129-269 FROM N.A.
RA Andrews S., Dubbelde C., Ryan E.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [7]
RP FUNCTION.
RX MEDLINE=92229472; PubMed=1373524;
RA Preston G.M., Carroll T.P., Guggino W.B., Agre P.;
RT "Appearance of water channels in Xenopus oocytes expressing red cell
RT CHIP28 protein.";
RL Science 256:385-387(1992).
RN [8]
RP TARGET OF MERCURY INHIBITION.
RX MEDLINE=93106996; PubMed=7677994;
RA Preston G.M., Jung J.S., Guggino W.B., Agre P.;
RT "The mercury-sensitive residue at cysteine 189 in the CHIP28 water
RT channel.";
RL J. Biol. Chem. 268:17-20(1993).
RN [9]
RP TOPOLOGY.
RX MEDLINE=94124503; PubMed=7507481;
RA Preston G.M., Jung J.S., Guggino W.B., Agre P.;
RT "Membrane topology of aquaporin CHIP. Analysis of functional epitope-
RT scanning mutants by vectorial proteolysis.";
RL J. Biol. Chem. 269:1668-1673(1994).
RN [10]
RP STRUCTURE BY ELECTRON CRYO-MICROSCOPY.
RX MEDLINE=94313979; PubMed=7518771;
RA Walz T., Smith B.L., Agre P., Engel A.;
RT "The three-dimensional structure of human erythrocyte aquaporin
RT CHIP.";
RL EMBO J. 13:2985-2993(1994).
RN [11]
RP STRUCTURE BY ELECTRON CRYO-MICROSCOPY.
RX MEDLINE=97320502; PubMed=9177353;
RA Walz T., Hirai T., Murata K., Heymann J.B., Mitsuoka K.,
RA Fujiyoshi Y., Smith B.L., Agre P., Engel A.;
RT "The three-dimensional structure of aquaporin-1.";
RL Nature 387:624-627(1997).
RN [12]
RP X-RAY CRYSTALLOGRAPHY (3.8 ANGSTROMS).
RX MEDLINE=20487015; PubMed=11034202;
RA Murata K., Mitsuoka K., Hirai T., Walz T., Agre P., Heymann J.B.,
RA Engel A., Fujiyoshi Y.;
RT "Structural determinants of water permeation through aquaporin-1.";
RL Nature 407:599-605(2000).
RN [13]
RP VARIANT BLOOD GROUP COLTON.
RX MEDLINE=94365170; PubMed=7521882;
RA Smith B.L., Preston G.M., Spring F., Anstee D.J., Agre P.;
RT "Human red cell aquaporin CHIP. I. Molecular characterization of ABH
RT and Colton blood group antigens.";
RL J. Clin. Invest. 94:1043-1049(1994).
RN [14]
RP FUNCTION: FORMS A WATER-SPECIFIC CHANNEL THAT PROVIDES THE PLASMA
RP MEMBRANES OF RED CELLS AND KIDNEY PROXIMAL TUBULES WITH HIGH
RP PERMEABILITY TO WATER, THEREBY PERMITTING WATER TO MOVE IN THE
RP DIRECTION OF AN OSMOTIC GRADIENT.
RC [1] SUBUNIT: HOMOTETRAMER.
CC [1] SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC [1] TISSUE SPECIFICITY: EXPRESSED IN A NUMBER OF TISSUES INCLUDING
CC ERYTHROCYTES, RENAL TUBULES, RETINAL PIGMENT EPITHELIUM,
CC HEART, LUNG, SKELETAL MUSCLE, KIDNEY AND PANCREAS. WEAKLY
CC EXPRESSED IN BRAIN, PLACENTA AND LIVER.
CC [1] DOMAIN: HB AND HE TOGETHER FORM A TRANSMEMBRANE HELIX. THEY ALSO
CC FUNCTION AS PORE HELICES INVOLVED IN DEFINING SUBSTRATE
CC SPECIFICITY. THE TWO NPA REPEATS ARE SITUATED AT THE END OF HB AND
CC HE AND ARE IMPORTANT FOR THE INTERACTION BETWEEN THESE TWO
CC HELICES.
CC [1] POLYMORPHISM: AQP1 IS RESPONSIBLE FOR THE COLTON BLOOD GROUP
CC SYSTEM. APPROXIMATELY 92% OF CAUCASIANS ARE CO(A+B-) (ALA-45),
CC APPROXIMATELY 8% ARE CO(A+B+), AND ONLY 0.2% ARE CO(A-B+) (VAL-
CC 45). CO(A-B-) WHICH IS VERY RARE, IS DUE TO A COMPLETE ABSENCE OF
CC AQP1.
CC [1] MISCELLANEOUS: PHARMACOLOGICALLY INHIBITED BY SUBMILLIMOLAR
CC CONCENTRATIONS OF Hg2+.
CC [1] SIMILARITY: BELONGS TO THE TRANSMEMBRANE CHANNEL MIP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M7829; AAA58425.1; -;
DR EMBL; U41517; AAC50648.1; -;
DR EMBL; U41518; AAC50649.1; -;
DR EMBL; S73482; AAB31193.1; -;
DR EMBL; AC004691; AAC16481.1; -;
DR EMBL; AC005155; AAC23788.1; -;
DR PIR; A41616; A41616.
DR PDB; IFQY; 18-OCT-00.
DR MIN; 107776; -;
DR MIN; 110450; -;
DR InterPro; IPR000425; -;
DR Pfam; PF00230; MIP; 1.
DR PRINTS; PR00783; MINTRINSCP.
DR PROSITE; PS00221; MIP; 1.
KW Transport; Transmembrane; Polymorphism; Blood group antigen;
KW Glycoprotein; 3D-structure.

```
FT DOMAIN 1 7 CYTOPLASMIC.
FT TRANSMEM 8 36 HELIX 1.
FT DOMAIN 37 48 EXTRACELLULAR.
FT TRANSMEM 49 66 HELIX 2.
FT DOMAIN 67 70 CYTOPLASMIC.
FT DOMAIN 71 76 IN MEMBRANE.
FT TRANSMEM 77 84 HELIX B.
FT DOMAIN 85 94 CYTOPLASMIC.
FT TRANSMEM 95 115 HELIX 3.
FT DOMAIN 116 136 EXTRACELLULAR.
FT TRANSMEM 137 155 HELIX 4.
FT DOMAIN 156 186 CYTOPLASMIC.
FT TRANSMEM 167 183 HELIX 5.
FT DOMAIN 184 186 EXTRACELLULAR.
FT TRANSMEM 187 192 IN THE MEMBRANE.
FT DOMAIN 193 200 HELIX E.
FT TRANSMEM 201 207 EXTRACELLULAR.
FT TRANSMEM 208 228 HELIX 6.
FT DOMAIN 229 269 CYTOPLASMIC.
FT REPEAT 76 78
FT REPEAT 192 194
FT DOMAIN 159 162
FT SITE 189 189
FT CARBOHYD 42 42
FT CARBOHYD 205 205
FT VARIANT 45 45
FT SEQUENCE 269 AA; 28526 MW; BA204D82FB26352E CRC64;

POLY-ARG.
HG(2+)-SENSITIVE RESIDUE.
N-LINKED (GLCNAC. . .).
N-LINKED (GLCNAC. . .) (POTENTIAL).
A -> V (IN CO(A-B+) ANTIGEN).
/FTIG-VAR_004400.

Query Match 19.2%; Score 51.5; DB 1; Length 269;
Best Local Similarity 26.1%; Pred. No. 18;
Matches 18; Conservative 13; Mismatches 17; Indels 21; Gaps 4;

QY 1 MELHISP--FKYP--HTQEAQKE-----AQRSLGMPGRHLGSSMSLAL- 41
Db 25 VFISIGSALGFKYPVGNQNTAVODNVKVSIAFGLSIATLAQSVGHISGAHLNPAVTLGLL 84

QY 42 --CLVPLVR 48
Db 85 LSCQISIFR 93
```

Search completed: July 30, 2001, 16:53:30
Job time: 190 sec

UD 89 LFSLRGEIVFPKNETAESEFEKSVHER--RQEGNAGSGRKQLDLDLVRRGH 137

Query Match 21.1%; Score 56.5; DB 2; Length 1061;
Best Local Similarity 31.0%; Pred. No. 28;
Matches 13; Conservative 11; Mismatches 17; Indels 1; Gaps 1;

RESULT	6
C75459	

Query Match 20.9%; Score 56; DB 2; Length 338;
Best Local Similarity 38.1%; Pred. No. 9.3;
Matches 8; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

lymphocyte-Peyer's patch adhesion molecule alpha 4 precursor - mouse
A411131
N:Alternate names: integrin alpha-4

n, map position: 3

n, map position: 3

n, map position: 3

Query Match 20.1%; Score 54; DB 2; Length 257;
Best Local Similarity 33.3%; Pred. No. 13;
Matches 16; Conservative 9; Mismatches 21; Indels

QY 12 PHTQEAQK-----EQRSLGEMPGRHLGSSMSIALCLVPLVREG 50
| | | | : | | | : | : | : | : | : |
Db 340 PETQVAQQLELAKGKLAEAAKQLG---GDLSFSFISFVSALLILLREG 384

Search completed: July 30, 2001, 16:52:13
Job time: 248 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 16:48:05 ; Search time 26.85 Seconds
(without alignments)
39.110 Million cell updates/sec

Title: US-09-402-713A-2
Perfect score: 268
Sequence: 1 MELHITSPKPKYPTQEAQRE.....HLGSSMSLALCLVPLVREGH 51

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	60.5	22.6	1239	2	US-08-937-931-2
2	60.5	22.6	1239	4	US-09-285-502-2
3	53	19.8	1151	1	US-08-286-889-37
4	53	19.8	1151	1	US-08-485-618-37
5	53	19.8	1151	1	US-08-362-652-37
6	53	19.8	1151	2	US-08-605-672-37
7	53	19.8	1151	2	US-08-482-293A-37
8	53	19.8	1151	2	US-08-943-363-37
9	53	19.8	1151	4	US-09-193-043-37
10	53	19.8	1161	1	US-08-485-618-55
11	53	19.8	1161	1	US-08-362-652-55
12	53	19.8	1161	1	US-08-605-672-55
13	53	19.8	1161	2	US-08-482-293A-55
14	53	19.8	1161	2	US-08-943-363-55
15	53	19.8	1161	4	US-09-193-043-55
16	52.5	19.6	685	3	US-09-031-563-21
17	52.5	19.6	1055	3	US-09-031-563-27
18	52.5	19.6	1315	3	US-09-031-563-2
19	52.5	19.6	1315	3	US-09-031-563-25
20	51.5	19.2	269	1	US-08-447-534-5
21	51.5	19.2	269	1	US-08-468-763-17
22	51.5	19.2	269	1	US-08-448-160-5
23	51.5	19.2	269	2	US-08-393-996A-17
24	51	19.0	1155	1	US-08-286-889-46
25	51	19.0	1155	1	US-08-485-618-46
26	51	19.0	1155	1	US-08-362-652-46
27	51	19.0	1155	2	US-08-605-672-46

28	51	19.0	1155	2	US-08-482-293A-46	Sequence 46, Appl
29	51	19.0	1155	2	US-08-943-363-46	Sequence 46, Appl
30	51	19.0	1155	4	US-09-193-043-46	Sequence 46, Appl
31	51	19.0	1161	1	US-08-173-497-2	Sequence 2, Appli
32	51	19.0	1161	1	US-08-286-889-2	Sequence 2, Appli
33	51	19.0	1161	1	US-08-485-618-2	Sequence 2, Appli
34	51	19.0	1161	1	US-08-485-618-53	Sequence 53, Appl
35	51	19.0	1161	1	US-08-485-618-59	Sequence 99, Appl
36	51	19.0	1161	1	US-08-362-652-2	Sequence 2, Appli
37	51	19.0	1161	1	US-08-362-652-53	Sequence 2, Appli
38	51	19.0	1161	2	US-08-605-672-2	Sequence 2, Appli
39	51	19.0	1161	2	US-08-605-672-53	Sequence 53, Appl
40	51	19.0	1161	2	US-08-605-672-99	Sequence 99, Appl
41	51	19.0	1161	2	US-08-482-293A-2	Sequence 2, Appli
42	51	19.0	1161	2	US-08-482-293A-53	Sequence 53, Appl
43	51	19.0	1161	2	US-08-482-293A-99	Sequence 99, Appl
44	51	19.0	1161	2	US-08-943-363-2	Sequence 2, Appli
45	51	19.0	1161	2	US-08-943-363-99	Sequence 99, Appl

ALIGNMENTS

RESULT 1
US-08-937-931-2
: Sequence 2, Application US/08937931
: Patent No. 5935792
: GENERAL INFORMATION:
: APPLICANT: Rubin, Gerald M.
: APPLICANT: Pan, DuoJia
: APPLICANT: Rooke, Jenny
: APPLICANT: Yavari, Reza
: APPLICANT: Xu, Tian
: TITLE OF INVENTION: KUZ: A No. 5935792el Family of Metalloproteases
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
: STREET: 268 BUSH STREET, SUITE 3200
: CITY: SAN FRANCISCO
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 94104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/937,931
: FILING DATE:
: CLASSIFICATION: 800
: ATTORNEY/AGENT INFORMATION:
: NAME: OSMAN, RICHARD A
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: B97-081
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 343-4341
: TELEFAX: (415) 343-4342
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1239 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
US-08-937-931-2

Query Match 22.6%; Score 60.5; DB 2; Length 1239;
Best Local Similarity 47.2%; Pred. No. 2.9;
Matches 17; Conservative 4; Mismatches 10; Indels 5; Gaps 2;
QY 9 FKYPHTQEAQKQSRSLGEM----PGRHLGSSMSLA 40

```
||||| |: ||| |: | |||| :||
Db 341 FKYPH-QKYTEANFAEGAFYDPTGRRLLGSSANVA 375

RESULT 2
US-09-285-502-2
; Sequence 2, Application US/09285502
; Patent No. 6190876
; GENERAL INFORMATION:
; APPLICANT: Rubin, Gerald M.
; APPLICANT: Pan, Duojia
; APPLICANT: Rooke, Jenny
; APPLICANT: Yavari, Reza
; APPLICANT: Xu, Tian
; TITLE OF INVENTION: KUZ: A No. 6190876el Family of Metalloproteases
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/285,502
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/937,931
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B97-081
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1239 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-285-502-2

Query Match 22.6%; Score 60.5; DB 4; Length 1239;
Best Local Similarity 47.2%; Pred. No. 2.9;
Matches 17; Conservative 4; Mismatches 10; Indels 5; Gaps 2;

QY 9 FKYPH-QKYTEANFAEGAFYDPTGRRLLGSSANVA 40
||||| |: ||| |: | |||| :||
Db 341 FKYPH-QKYTEANFAEGAFYDPTGRRLLGSSANVA 375

RESULT 3
US-08-286-889-37
; Sequence 37, Application US/08286889
; Patent No. 5470953
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Mich
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5470953el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
```

```
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/286,889
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: P38,659
; REFERENCE/DOCKET NUMBER: 27866/32168
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-286-889-37

Query Match 19.8%; Score 53; DB 1; Length 1151;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

QY 14 TQEAQKEAQRSLGMPGRHLGSSMSLALCLVPLVREG 50
||||| |: ||| |: | |||| :||
Db 428 TQEARHWRPKS--EVRGTQIGSYFGASLCSDVDVDRDG 462

RESULT 4
US-08-485-618-37
; Sequence 37, Application US/08485618
; Patent No. 5728533
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,618
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: P38,659
; REFERENCE/DOCKET NUMBER: 27866/32168
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-286-889-37
```

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/362,652
;; FILING DATE: 21-DEC-1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Williams Jr., Joseph A.
;; REGISTRATION NUMBER: 38,659
;; REFERENCE/DOCKET NUMBER: 27866/32797
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 312-474-6300
;; TELEFAX: 312-474-0448
;; TELEX: 25-3856
;; INFORMATION FOR SEQ ID NO: 37:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1151 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-485-618-37

Query Match 19.8%; Score 53; DB 1; Length 1151;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

Qy 14 TQEAQKEAQRSLGEMPGRHLSGSSMSLALCLVPLVREG 50
||||: :| | :| | :| | :| | :| | :| | :| |
Db 428 TQEARHWRPKS--EVRGTQIGSYFGASLCSVDVDRDG 462

RESULT 5
US-08-362-652-37
; Sequence 37, Application US/08362652
; Patent No. 5766850
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vliet, Monica
; TITLE OF INVENTION: NO. 5766850el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,652
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32391
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

;; MOLECULE TYPE: protein
US-08-362-652-37

Query Match 19.8%; Score 53; DB 1; Length 1151;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

Qy 14 TQEAQKEAQRSLGEMPGRHLSGSSMSLALCLVPLVREG 50
||||: :| | :| | :| | :| | :| | :| | :| |
Db 428 TQEARHWRPKS--EVRGTQIGSYFGASLCSVDVDRDG 462

RESULT 6
US-08-605-672-37
; Sequence 37, Application US/08605672
; Patent No. 5817515
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vliet, Monica
; TITLE OF INVENTION: NO. 5817515el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/605,672
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32684
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-605-672-37

Query Match 19.8%; Score 53; DB 2; Length 1151;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

Qy 14 TQEAQKEAQRSLGEMPGRHLSGSSMSLALCLVPLVREG 50
||||: :| | :| | :| | :| | :| | :| | :| |
Db 428 TQEARHWRPKS--EVRGTQIGSYFGASLCSVDVDRDG 462

```

STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/943,363
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32684
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 1151 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-943-363-37

Query Match 19.8% Score 53; DB 2; Length 1151;
Best Local Similarity 37.88; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels

QY 14 TQEAQKFAQRSLGEMPGRHGSSMLALCLVPLVREG 50
      |||| :| | | :|| :||| :| |
Db 428 TQEARHWRPKS-EVRGTQIGSYFGASLCSVDVDRDG 462

RESULT 9
US-09-193-043-37
; Sequence 37, Application US/09193043
; Patent No. 6251395
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; TITLE OF INVENTION: No. 6251395el Human 2
; FILE REFERENCE: 27866/35004
; CURRENT APPLICATION NUMBER: US/09/193.043
; CURRENT FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: 08/173,497
; EARLIER FILING DATE: 1993-12-23
; EARLIER APPLICATION NUMBER: 08/286,889
; EARLIER FILING DATE: 1994-08-05
; EARLIER APPLICATION NUMBER: 08/362,652
; EARLIER FILING DATE: 1994-12-21
; EARLIER APPLICATION NUMBER: 08/943,363
; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
; LENGTH: 1151
; TYPE: PRT
; ORGANISM: Rattus rattus
US-09-193-043-37

```



```

Query Match          19.8%; Score 53; DB 4; Length 1151;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

Qy 14 TOEAQKQSRSLGEMPGRHLSMSLALCLVPLVREG 50
      ||||: :| | :| | :| | :| | :| |
Db 428 TQEARHWRPKS--EVRGTQIGSYFGASLCSDVDVDRDG 462

RESULT 10
US-08-485-618-55
; Sequence 55, Application US/08485618
; Patent No. 5728533
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485.618
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32797
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-0448
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-485-618-55

Query Match          19.8%; Score 53; DB 1; Length 1161;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

Qy 14 TOEAQKQSRSLGEMPGRHLSMSLALCLVPLVREG 50
      ||||: :| | :| | :| | :| | :| |
Db 438 TQEARHWRPKS--EVRGTQIGSYFGASLCSDVDVDRDG 472

RESULT 11
US-08-362-652-55
; Sequence 55, Application US/08485618
; Patent No. 5728533
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

; Sequence 55, Application US/08362652
; Patent No. 5766850
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362.652
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32391
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-362-652-55

Query Match          19.8%; Score 53; DB 1; Length 1161;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

Qy 14 TOEAQKQSRSLGEMPGRHLSMSLALCLVPLVREG 50
      ||||: :| | :| | :| | :| | :| |
Db 438 TQEARHWRPKS--EVRGTQIGSYFGASLCSDVDVDRDG 472

RESULT 12
US-08-605-672-55
; Sequence 55, Application US/08605672
; Patent No. 5817515
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

Query Match 19.8%; Score 53; DB 2; Length 1161;
Best Local Similarity 37.8%; Pred. NO. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

RESULT 14
US-08-943-363-55
; Sequence 55, Application US/08943363
; Patent No. 5837478

/ AFFICCIANI: van der Vleuten, Monica
 / TITLE OF INVENTION: No. 5837478el Human 2 Integrin Alpha Subunit
 / NUMBER OF SEQUENCES: 114
 / CORRESPONDENCE ADDRESS:
 / ADDRESSSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 / STREET: 233 South Wacker Drive, 6300 Sear Tower
 /

CITY: Chicago
 STATE: Illinois
 COUNTRY: United States
 ZIP: 60606-6402
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/943,363
 FILING DATE:
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/173,497
 FILING DATE: 23-DEC-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/286,889
 FILING DATE: 5-AUG-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/362,652
 FILING DATE: 21-DEC-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Williams Jr., Joseph A.
 REGISTRATION NUMBER: 38,659
 REFERENCE/DOCKET NUMBER: 27866/32684
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 312-474-6300
 TELEFAX: 312-474-0448
 TELEX: 25-3856
 INFORMATION FOR SEQ ID NO: 55:
 SEQUENCE CHARACTERISTICS:

;
; LENGTH: 1161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-943-363-55

Query Match 19.8%; Score 53; DB 2; Length 1161;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;
Qy 14 TOEAQKEAQRSLGEMPGRHLSGSSMSLALCLVPLVREG 50
||||: :| |:| :||| :||| :|:
Db 438 TOEARHWRPKS--EVRGTQIGSYFGASLCSVDVDRDG 472

RESULT 15
US-09-193-043-55
; Sequence 55, Application US/09193043
; Patent No. 6251395
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: NO. 6251395el Human 2
; FILE REFERENCE: 27866/35004
; CURRENT APPLICATION NUMBER: US/09/193,043
; CURRENT FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: 08/173,497
; EARLIER FILING DATE: 1993-12-23
; EARLIER APPLICATION NUMBER: 08/286,889
; EARLIER FILING DATE: 1994-08-05
; EARLIER APPLICATION NUMBER: 08/362,652
; EARLIER FILING DATE: 1994-12-21
; EARLIER APPLICATION NUMBER: 08/943,363
; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 55
; LENGTH: 1161
; TYPE: PRT
; ORGANISM: Rattus rattus
US-09-193-043-55

Query Match 19.8%; Score 53; DB 4; Length 1161;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;
Qy 14 TOEAQKEAQRSLGEMPGRHLSGSSMSLALCLVPLVREG 50
||||: :| |:| :||| :||| :|:
Db 438 TOEARHWRPKS--EVRGTQIGSYFGASLCSVDVDRDG 472

Search completed: July 30, 2001, 16:50:50
Job time: 165 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 16:48:05 ; Search time 42.83 Seconds
(without alignments)
72.188 Million cell updates/sec

Title: US-09-402-713A-2
Perfect score: 268
Sequence: 1 MFLHSSPKPKYPTQEAQKE.....HLGSSMSLALCLVPLVREGH 51

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues
Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0601.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	268	100.0	51	19 AAW79736	Prostate cancer an
2	268	100.0	51	19 AAW79738	Prostate cancer an
3	77	28.7	16	20 AAY07887	Human secreted pro
4	63	23.5	514	21 AAY87343	Human signal pepti
5	63	23.5	602	21 AAB43919	Human cancer assoc
6	60.5	22.6	1239	19 AAW56131	Drosophila melanoga
7	59	22.0	897	20 AAY07092	Colon cancer assoc
8	59	22.0	1141	21 AAB43008	Human ORF ORF2772
9	56	20.9	11	19 AAW79737	Prostate cancer an
10	55	20.5	184	21 AAB40271	Human ORF ORF35 p
11	55	20.5	2432	21 AAY85565	Human homologue of

12	54.5	20.3	99	20	AAW29863	Human secreted pro
13	54.5	20.3	100	20	AAW25713	Human secreted pro
14	54.5	20.3	461	17	AAW98341	Human p57 protein.
15	54.5	20.3	461	17	AAW98342	Bovine p57 protein
16	54.5	20.3	598	17	AAW98344	GST/truncated huma
17	54.5	20.3	673	17	AAW98343	GST/human p57 fusi
18	54	20.1	414	21	AAG22216	Arabidopsis thalia
19	54	20.1	414	21	AAG46290	Arabidopsis thalia
20	54	20.1	450	21	AAG22215	Arabidopsis thalia
21	54	20.1	450	21	AAG46289	Arabidopsis thalia
22	54	20.1	453	21	AAG22214	Arabidopsis thalia
23	54	20.1	453	21	AAG46288	Arabidopsis thalia
24	53.5	20.0	112	20	AAW60194	Human endometrium
25	53	19.8	1151	16	AAW78179	Rat alpha-d compos
26	53	19.8	1151	16	AAW23059	Rat beta 2 integri
27	53	19.8	1151	19	AAW72834	Rat alpha-d #2. R
28	53	19.8	1151	19	AAW5101	Rat beta-integrin
29	53	19.8	1151	19	AAW60001	Rat alpha d polype
30	53	19.8	1151	20	AAW73344	Rat alphas protein
31	53	19.8	1151	21	AAW07371	Rat alpha_d partia
32	53	19.8	1161	16	AAW78169	Rat alpha-d subuni
33	53	19.8	1161	18	AAW23062	Rat beta 2 integri
34	53	19.8	1161	19	AAW72824	Rat alpha-d #1. R
35	53	19.8	1161	19	AAW65104	Rat beta-integrin
36	53	19.8	1161	19	AAW60004	Rat alpha d polype
37	53	19.8	1161	20	AAW73345	Rat alphas protein
38	53	19.8	1161	21	AAW07374	Rat alpha_d protei
39	52.5	19.6	109	20	AAW48422	Human prostate can
40	52.5	19.6	1055	20	AAW31730	Human fused protei
41	52.5	19.6	1315	20	AAW43265	Human Fused Relate
42	52.5	19.6	1315	20	AAW31725	Human fused protei
43	52.5	19.6	1315	20	AAW31729	Human fused protei
44	52.5	19.6	1315	21	AAW30679	Amino acid sequenc
45	52.5	19.6	1333	20	AAW31727	Human fused CDNA i

ALIGNMENTS

RESULT 1
AAW79736
ID AAW79736 standard; Protein; 51 AA.
XX
AC AAW79736;
XX
DT 30-DEC-1998 (first entry)
XX
DE Prostate cancer antigen (PCA3) protein variant 1.
XX
KW Prostate cancer antigen protein variant 1; PCA3; prostatic cancer;
PC.
XX
OS Homo sapiens.
XX
PN WO9845420-Al.
XX
PD 15-OCT-1998.
XX
PF 09-APR-1998; 98WO-CA00346.
XX
PR 10-APR-1997; 97US-0041836.
XX
PA (DIAG-) DIAGNOCURE INC.
XX
PI Bussemakers MJG;
XX
DR WPI; 1998-568347/48.
XX
N-PSDB; AAW62427.
XX
PT New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
PT prevention and treatment of prostatic cancer
XX
PS Claim 16; Fig 2B-2J; 111pp; English.

XX The present sequence represents the prostate cancer antigen (PCA3).
 CC protein variant 1 encoded by a PCA3 spliced cDNA sequence comprising
 CC of exons 1, 2, 3, 4a and 4b of the PCA3 gene. The invention claims
 CC for PCA3 cDNA variants and the proteins they encode. The invention
 CC also claims for antibodies against PCA3 protein. The antibodies are
 CC claimed to be useful for detecting PCA3 protein in immunoassay tests,
 CC for diagnosing, assessing and prognosing of prostatic cancer (PC).
 CC Antibodies, optionally coupled to a cytotoxin or radioisotope, and
 CC nucleic acids antisense to PCA3 cDNA are claimed to be useful for
 CC treating PC, while determining elevated levels of PCA3 (as RNA or
 CC protein) is useful for detecting a predisposition to development of
 CC PC, e.g. in prenatal tests. Detecting PCA3 protein allows
 CC differentiation between malignant and benign prostatic disease,
 CC and the level of PCA3 expression allows correlation with the grade of
 CC tumour. PCA3 protein and its fragments are also claimed to be useful
 CC in vaccines for preventing PC; in drug screens for identifying
 CC specific (antagonists (potentially useful therapeutically) and for
 CC studying protein-DNA interactions.
 XX
 SQ Sequence 51 AA;

Query Match 100.0%; Score 268; DB 19; Length 51;
 Best Local Similarity 100.0%; Pred. NO. 4.7e-32;
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFLHSSPFKYPHTQEAQKRAQSLGEMPGRHGSSMSLALCLVPLVREGH 51
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 mflhsspfkyphtqeaqkraqslgempgrhlgssmslalcivplvregh 51

RESULT 2

AAW79738
 ID AAW79738 standard; Protein; 51 AA.

AC AAW79738;

XX 30-DEC-1998 (first entry)

XX Prostate cancer antigen (PCA3) wild-type protein.

XX Prostate cancer antigen protein variant 1; PCA3; prostatic cancer;
 KW PC.

XX Homo sapiens.

XX WO9845420-A1.

XX 15-OCT-1998.

XX 09-APR-1998; 98WO-CA00346.

XX 10-APR-1997; 97US-0041836.

XX (DIAG-) DIAGNOCURE INC.

XX Bussemakers MJG;

XX WPI; 1998-568347/48.

XX N-PSDB; AAV62430.

XX New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
 PT prevention and treatment of prostatic cancer

XX Claim 16; Fig 5B-5F; 111pp; English.

PS The present sequence represents the prostate cancer antigen (PCA3)
 XX wild-type protein sequence encoded by a PCA3 wild-type cDNA sequence
 CC comprising of exons 1, 2, 3, 4a-4d of the PCA3 gene. The invention
 CC claims for PCA3 cDNA variants and the proteins they encode. The
 CC invention also claims for antibodies against PCA3 protein. The
 CC antibodies are claimed to be useful for detecting PCA3 protein in

CC immunoassay tests, for diagnosing, assessing and prognosing of
 CC prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin
 CC or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed
 CC to be useful for treating PC, while determining elevated levels of
 CC PCA3 (as RNA or protein) is useful for detecting a predisposition to
 CC development of PC, e.g. in prenatal tests. Detecting PCA3 protein
 CC allows differentiation between malignant and benign prostatic disease,
 CC and the level of PCA3 expression allows correlation with the grade of
 CC tumour. PCA3 protein and its fragments are also claimed to be useful
 CC in vaccines for preventing PC; in drug screens for identifying
 CC specific (antagonists (potentially useful therapeutically) and for
 CC studying protein-DNA interactions.
 XX
 SQ Sequence 51 AA;

Query Match 100.0%; Score 268; DB 19; Length 51;
 Best Local Similarity 100.0%; Pred. NO. 4.7e-32;
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFLHSSPFKYPHTQEAQKRAQSLGEMPGRHGSSMSLALCLVPLVREGH 51
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 mflhsspfkyphtqeaqkraqslgempgrhlgssmslalcivplvregh 51

RESULT 3

AAV07887
 ID AAV07887 standard; Protein; 16 AA.

XX AAV07887;

XX 06-JUL-1999 (first entry)

XX Human secreted protein fragment encoded from gene 36.

XX Human; secreted protein; treatment; prevention; protein therapy; AIDS;
 KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
 KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
 KW immune system disease; autoimmune disease; hepatic disease; lymphoma;
 KW renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
 KW cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
 KW pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
 KW arthritis; malignancy; digestive; endocrine; infection.

XX Homo sapiens.

XX WO9918208-A1.

XX 15-APR-1999.

XX 01-OCT-1998; 98WO-US20775.

XX 02-OCT-1997; 97US-0060884.

XX 02-OCT-1997; 97US-0060833.

XX 02-OCT-1997; 97US-0060836.

XX 02-OCT-1997; 97US-0060837.

XX 02-OCT-1997; 97US-0060838.

XX 02-OCT-1997; 97US-0060839.

XX 02-OCT-1997; 97US-0060843.

XX 02-OCT-1997; 97US-0060862.

XX 02-OCT-1997; 97US-0060866.

XX 02-OCT-1997; 97US-0060874.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Carter KC, Duan DR, Endress GA, Feng P, Ferrie AM;
 PI Florence KA, Greene JM, Janat F, Lafleur DW, Ni J;
 PI Rosen CA, Ruben SM, Shi Y, Young P, Yu G;
 XX WPI; 1999-264022/22.
 DR N-PSDB; AAX37486.

PT New isolated human genes and the secreted polypeptides they encode

XX PS Claim 1b: Page 297; 368pp; English.
 XX CC This invention describes novel isolated human genes and the secreted
 CC proteins they encode. The products of the invention are useful for
 CC preventing, treating or ameliorating medical conditions, e.g. by protein
 CC or gene therapy. Also pathological conditions can be diagnosed by
 CC determining the amount of the new polypeptides in a sample or by
 CC determining the presence of mutations in the 101 polynucleotides.
 CC Specific uses are described for each of the 101 polynucleotides, based
 CC on which tissues they are most highly expressed in, and include
 CC neurodegenerative disorders, developmental abnormalities and fetal
 CC deficiencies, blood disorders, leukemias, diseases of the immune system,
 CC autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,
 CC allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate
 CC disease, skeletal or cardiac muscle disorders, pulmonary disorders,
 CC transplant rejection, disorders involving osteoclasts such as
 CC osteoporosis, arthritis or malignancies, digestive/endocrine disorders,
 CC infections and AIDS. The human secreted proteins of the invention are
 CC represented in AAY07852-Y07993 and the encoding nucleic acids are
 CC represented in AAX37451-X37552.

XX SQ Sequence 16 AA;

Query Match 28.7%; Score 77; DB 20; Length 16;
 Best Local Similarity 100.0%; Pred. No. 0.00012; Mismatches 0; Indels 0; Gaps 0;
 Matches 15; Conservative 0;

QY 37 MSLALCLVPLVREGH 51
 Db 1 mslalclvplvregh 15

RESULT 4
 AAY87343
 ID AAY87343 standard; Protein; 514 AA.
 AC AAY87343;
 XX

DT 11-MAY-2000 (first entry)

DE Human signal peptide containing protein HSP-120 SEQ ID NO:120.

KW Human; signal peptide-containing protein; HSP; diagnosis: cancer;
 KW inflammation; cardiovascular disease; anticancer; anti-inflammatory;
 KW antimicrobial; nontropic; neuroprotective; cardiovascular; hepatotropic;
 KW antiasthmatic; gene therapy; cell proliferation; neurological disorder;
 KW reproductive disorder; developmental disorder; arteriosclerosis;
 KW cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;
 KW asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;
 KW Parkinson's disease; Huntington's diseases; ovulatory defect;

XX muscular dystrophy.

XX Homo sapiens.

XX WO200000610-A2.

XX PD 06-JAN-2000.

XX PF 25-JUN-1999; 99WO-US14484.

XX PR 26-JUN-1998; 98US-0090762.

XX PR 31-JUL-1998; 98US-0094983.

XX PR 01-OCT-1998; 98US-0102686.

XX PR 11-DEC-1998; 98US-0112129.

XX (INCY-) INCYTE PHARM INC.

XX Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KJ, Baughn MR;

PI Akerblom IE, Au-Young J, Yue H, Patterson C, Reddy R, Hillman JL;
 PI Bandman O;

XX WPI; 2000-160673/14.
 DR N-PSDB; AAZ98228.
 XX PT New human signal peptide-containing proteins useful in treatment,
 PT prevention and diagnosis of e.g. cancer, inflammation and
 PT cardiovascular disease
 XX Claim 1; Page 239-240; 327pp; English.
 XX AAY98109 to AAZ98242 encode AAY87224 to AAY87357 which represent the
 CC human signal peptide-containing proteins HSP-1 to HSP-134. HSPs have
 CC anticancer, anti-inflammatory, antimicrobial, nontropic, hepatotropic,
 CC neuroprotective, cardiovascular and antiasthmatic activities, and can
 CC be used in gene therapy. HSPs can be used to treat or prevent disorders
 CC associated with decreased activity or function of HSP. Antagonists of
 CC HSP are used to treat or prevent disorders associated with increased
 CC activity or function of HSP. Such diseases include cell proliferation
 CC (including cancer), inflammation, cardiovascular, neurological,
 CC reproductive or developmental disorders, (e.g. arteriosclerosis,
 CC cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,
 CC asthma, Crohn's disease, microbial or other infections, congestive or
 CC ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's
 CC diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSP
 CC nucleic acids can be used for the recombinant production of HSP, for
 CC detecting HSP in standard hybridisation and amplification assays (for
 CC diagnosis and monitoring), in gene therapy, as antisense,
 CC triplex-forming or ribozyme therapeutics, for detecting related sequences
 CC or genetic variations, and for chromosomal mapping. HSP are also used to
 CC raise specific antibodies (Ab) and to screen for agonists and
 CC antagonists (potential therapeutic agents). Ab are used to diagnose, or
 CC monitor, HSP-related diseases (in usual immunoassays), as therapeutic
 CC antagonists, in competitive drug screens, and for purification of HSP
 CC from natural sources.

XX SQ Sequence 514 AA;

Query Match 23.5%; Score 63; DB 21; Length 514;
 Best Local Similarity 42.1%; Pred. No. 0.92; Mismatches 7; Indels 2; Gaps 2;
 Matches 16; Conservative 7;

QY 12 PHTQPAQKQARSL-GEMPGRHLSMSLAL-CLVPLV 47
 Db 471 praedlaedsgslygrapgrhtwlllaalacvpll 508

RESULT 5
 AAB43919

ID AAB43919 standard; Protein; 602 AA.

XX AC AAB43919;

XX DT 08-FEB-2001 (first entry)

XX DE Human cancer associated protein sequence SEQ ID NO:1364.

XX Human; cancer associated gene; cancer antigen; detection: cancer;
 KW diagnosis; cytostatic; proliferative; vulnery; immunomodulator;
 KW antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
 KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
 KW dermatological; neuroprotective; thrombolytic; coagulant; nontropic;
 KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
 KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
 KW allergic reaction; graft versus host disease; organ rejection;
 KW haemostatic; thrombolytic; cardiovascular disorder; infection;
 KW neurological disease; drug screening.

OS Homo sapiens.

XX WO200055350-A1.

XX 21-SEP-2000.

PD

XX	08-MAR-2000; 2000WO-US05882.
PF	12-MAR-1999; 99US-0124270.
XX	(HUMA-) HUMAN GENOME SCI INC.
XX	Rosen CA, Ruben SM;
XX	WPI: 2000-587533/55.
DR	N-PSDB; AAC78128.
DR	
XX	Novel isolated nucleic acids comprising sequences encoding peptides
PT	useful for treating or diagnosing e.g. cancer -
PT	Claim 11; Page 2024-2026; 2352pp; English.
XX	
CC	AAC77607 to AAC78448 encode the human cancer associated proteins given
CC	in AAB43398 to AAB44239. The proteins can have activities based on the
CC	tissues and cells the genes are expressed in. Example of activities
CC	include: cytostatic; proliferative; vulnerrary; immunomodulator;
CC	antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC	antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC	dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC	neotropic; vasotropic; antipsoriatic and antiangiogenic. The
CC	polynucleotides and polypeptides can be used for preventing, treating or
CC	ameliorating medical conditions and diagnosing pathological conditions.
CC	Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC	the present invention may be used to treat immune disorders by activating
CC	or inhibiting the proliferation, differentiation or mobilisation of
CC	immune cells, to treat disorders of haematopoietic cells, autoimmune
CC	disorders, allergic reactions, graft versus host disease and organ
CC	rejection, modulate haemostatic or thrombolytic activity, modulate
CC	inflammation, cancers, cardiovascular disorders, neurological disease and
CC	bacterial or viral infections. The peptides, nucleotides, antibodies, and
CC	agonists and antagonists may be also be used in drug screens. AAC78449 to
CC	AAC78457 and AAB44240 represent sequences used in the exemplification of
CC	the present invention.
XX	
SQ	Sequence 602 AA;
	Query Match 23.5%; Score 63; DB 21; Length 602;
	Best Local Similarity 42.1%; Pred. No. 1.1;
	Matches 16; Conservative 7; Mismatches 13; Indels 2; Gaps 2;
QY	12 PHTQEAQKQARSL-CEMPGRHLGSSMSIAL-CLVPLV 47
	: : : : : :
Db	559 praedlaedsgslygrabgrhtwslilaalacivpll 596
RESULT	6
AAW56131	
ID	AAW56131 standard; Protein; 1239 AA.
XX	
XX	AAW56131;
XX	
DT	17-AUG-1998 (first entry)
XX	
DE	Drosophila melanogaster KUZ protein.
XX	
KW	kuzbanian; kuz; neurogenic; KUZ protein; neuronal partitioning;
KW	development; NOTCH protein processing; regulation; cell function;
KW	signal transduction pathways; screening; receptor binding;
KW	metalloprotease.
XX	
OS	Drosophila melanogaster.
XX	
PN	WO9808933-Al.
XX	
PD	05-MAR-1998.
XX	
PF	27-AUG-1997; 97WO-US15099.

DR WPI; 1999-132448/11.

XX New isolated cancer associated nucleic acids and polypeptides -

PT isolated using sera from cancer patients, used to develop products

PT for the diagnosis, monitoring or treatment of cancers

XX

XX Disclosure; Page 652-654; 787pp; English.

PS

CC The invention relates to a method for diagnosing a disorder characterised

CC by expression of a human cancer associated antigen precursor coded for by

CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a

CC biological sample isolated from a subject with an agent that specifically

CC binds to the NAM, an expression product or a fragment of an expression

CC product complexed with an HLA molecule; and (b) determining the

CC interaction between the agent and the NAM or the expression product as a

CC determination of the disorder. The products and methods can be used in

CC the diagnosis, monitoring, research, or treatment of conditions

CC characterised by the expression of various cancer associated antigens.

CC The invention provides nucleic acid sequences and encoded polypeptides

CC which are cancer associated antigen precursors expressed in human breast

CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and

CC lung cancer.

XX

SQ Sequence 897 AA;

Query Match 22.0%; Score 59; DB 20; Length 897;

Best Local Similarity 32.0%; Pred. No. 6.9;

Matches 16; Conservative 14; Mismatches 16; Indels 4; Gaps 2;

QY 4 HISSPFKYPHTQEAQKEAQRSL---GEMPGRHLSGSSMLALCLVPLVREG 50

Db 198 hitaspkistqgeaerqalsirggtltgkfmsts-sipgcllgvaleg 246

RESULT 8

AAAB43008

ID AAB43008 standard; Protein; 1141 AA.

XX

AC AAB43008;

XX

DT 08-FEB-2001 (first entry)

XX

DE Human ORFX ORF2772 polypeptide sequence SEQ ID NO:5544.

XX

KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;

KW vulnerable; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;

KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;

KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;

KW hypotensive; dermatological; immunosuppressive; antiinflammatory;

KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;

KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;

KW neurodegenerative disorder; osteoarthritis; graft vs host disease;

KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;

KW cholesterol ester storage; systemic lupus erythematosus; infection;

KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;

KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;

KW bone damage; cartilage damage; antiinflammatory disease; coagulation;

KW thrombosis; contraceptive.

XX

OS Homo sapiens.

XX

PN WO200058473-A2.

XX

PD 05-OCT-2000.

XX

PF 31-MAR-2000; 2000WO-US08621.

XX

PR 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

XX

PA (CURA-) CURAGEN CORP.

XX

PI Shimkets RA, Leach M;

XX

XX WPI; 2000-602362/57.

DR N-PSDB; AAC77217.

XX

PT Novel nucleic acids and peptides derived from open reading frame X,

PT useful for treating e.g. cancers, proliferative disorders,

PT neurodegenerative disorders and cardiovascular disease -

XX

PS Claim 11; Page 4728-4731; 5507pp; English.

XX

CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,

CC which represent the human ORFX open reading frames 1 to 3161. The ORFX

CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;

CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;

CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;

CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;

CC antidiabetic; hypotensive; dermatological; immunosuppressive;

CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;

CC antithyroid; and antianemic. The sequences can be used for determining

CC the presence of or predisposition to, or preventing or treating

CC pathological conditions associated with an ORFX-associated disorder. The

CC nucleic acids can be used to express ORFX proteins in gene therapy

CC vectors. The proteins and nucleic acids may be used to treat cancers,

CC proliferative disorders, neurodegenerative disorders, osteoarthritis,

CC graft vs host disease, cardiovascular disease, diabetes mellitus,

CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus

CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,

CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,

CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,

CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance

CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX

SQ Sequence 1141 AA;

Query Match 22.0%; Score 59; DB 21; Length 1141;

Best Local Similarity 32.0%; Pred. No. 9.3;

Matches 16; Conservative 14; Mismatches 16; Indels 4; Gaps 2;

QY 4 HISSPFKYPHTQEAQKEAQRSL---GEMPGRHLSGSSMLALCLVPLVREG 50

Db 405 hitaspkistqgeaerqalsirggtltgkfmsts-sipgcllgvaleg 453

RESULT 9

AAW79737

ID AAW79737 standard; Peptide; 11 AA.

XX

AC AAW79737;

XX

XX 30-DEC-1998 (first entry)

DT

DE Prostate cancer antigen (PCA3) protein antigenic peptide.

XX

XX Prostate cancer antigen cDNA splice variant 1; PCA3; prostatic cancer;

KW PC; PCA3 antigenic peptide.

XX

OS Homo sapiens.

XX

PN WO9845420-A1.

XX

PD 15-OCT-1998.

XX

PF 09-APR-1998; 98WO-CA00346.

XX

PR 10-APR-1997; 97US-0041836.

XX

PA (DIAG-) DIAGNOCURE INC.

XX

PI Bussemakers MJG;

XX DR WPI; 1998-568347/48.
 XX PT New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
 XX PT prevention and treatment of prostatic cancer
 XX PS Disclosure; Page 21; 111pp; English.
 XX CC The present sequence represents an antigenic peptide fragment
 CC derived from prostate cancer antigen (PCA3) protein. The invention
 CC claims for PCA3 cDNA variants and the proteins they encode. The
 CC invention also claims for antibodies against PCA3 protein. The
 CC antibodies are claimed to be useful for detecting PCA3 protein in
 CC immunoassay tests, for diagnosing, assessing and prognosing of
 CC prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin
 CC or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed
 CC to be useful for treating PC, while determining elevated levels of
 CC PCA3 (as RNA or protein) is useful for detecting a predisposition
 CC to development of PC, e.g. in prenatal tests. Detecting PCA3 protein
 CC allows differentiation between malignant and benign prostatic disease,
 CC and the level of PCA3 expression allows correlation with the grade of
 CC tumour. PCA3 protein and its fragments are also claimed to be useful
 CC in vaccines for preventing PC; in drug screens for identifying
 CC specific (ant)agonists (potentially useful therapeutically) and for
 CC studying protein-DNA interactions.
 XX SQ Sequence 11 AA;
 Query Match 20.9%; Score 56; DB 19; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.093;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 13 HTOEAOKEAQR 23
 Db 1 htgeaqkeagr 11
 |||||
 RESULT 10
 AAB40271
 ID AAB40271 standard; Protein; 184 AA.
 XX AC AAB40271;
 XX DT 08-FEB-2001 (first entry)
 XX DE Human ORFX ORF35 polypeptide sequence SEQ ID NO:70.
 XX KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX OS Homo sapiens.
 XX PN WO200058473-A2.
 XX PD 05-OCT-2000.
 XX PF 31-MAR-2000; 2000WO-US08321.
 XX PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.
 XX 30-MAR-2000; 2000US-0540763.
 PA (CURA-) CURAGEN CORP.
 XX PI Shimkets RA, Leach M;
 XX DR WPI; 2000-602362/57.
 DR N-PSDB; AAC74480.
 XX PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX PS Claim 11; Page 455; 5507pp; English.
 XX CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antianaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX SQ Sequence 184 AA;
 Query Match 20.5%; Score 55; DB 21; Length 184;
 Best Local Similarity 34.0%; Pred. No. 3.9;
 Matches 16; Conservative 3; Mismatches 16; Indels 12; Gaps 1;
 QY 6 SSPKYPHTQEAQKEAQRSLG-----EMPGRHLGSSMSLA 40
 Db 77 smpgksppapskegersrsklsnglpqkqpqldgrhsssssa 123
 |||||
 RESULT 11
 AAY85565
 ID AAY85565 standard; Protein; 2432 AA.
 XX AC AAY85565;
 XX DT 07-JUL-2000 (first entry)
 XX DE Human homologue of UNC-53 (Hs-UNC-53/2) sequence.
 XX KW UNC-53; Caenorhabditis elegans; microtubule; neural regeneration;
 KW anticancer; anti-neurodegeneration; antifibrotic; anti-adhesive; human;
 KW antisclerotic; antimetastatic; anti-arthritis; autoimmune disease.
 XX OS Homo sapiens.
 XX FH Key Location/Qualifiers
 FT Region 1..89
 FT /note= "this region can be replaced with one of the
 FT three sequences shown in AAY85566 to AAY85568;
 FT this creates three variants at the N-terminus"
 FT Misc-difference 1018
 FT /label= Asp or Glu

```

FT Region 1776..1778
FT /note= "present or absent depending upon the allele
FT /note= "from which the protein is translated"
XX
XX WO9963080-A1.
XX
XX 09-DEC-1999.
XX
XX 02-JUN-1999; 99WO-EP03848.
XX
XX 03-JUN-1998; 98GB-0011962.
XX
XX (JANC ) JANSSEN PHARM NV.
XX
XX Luyten WHML, De Raeymaekers MC, Geysen JJGH, Bogaert TAOE;
XX Maerten LJS, Verhasselt P, Van De Craen M;
XX
XX WPI: 2000-116370/10.
XX N-PSDB; AAA07836.
XX
XX Novel proteins and nucleic acids e.g. for treating neurodegeneration
XX
XX Claim 95; Fig 1d; 146pp; English.
XX
XX The invention provides vertebrate (human) protein homologue of a UNC-53
XX protein of Caenorhabditis elegans. The UNC-53 binds to microtubules or
XX their plus ends. The UNC-53 sequences are used to promote neural
XX regeneration, revascularization and wound healing; also for treating
XX neurodegenerative diseases, acute traumatic injury, fibrotic disease and
XX autoimmune diseases (e.g. rheumatoid arthritis and sclerosis). The UNC-53
XX polynucleotides can be used for recombinant production of the proteins,
XX as a source of probes for detecting allelic variants and polymorphisms,
XX for sequencing genomic DNA and for detecting UNC-53 expression; and as
XX source of therapeutic antisense sequences. Cells that express the
XX protein are used to identify regulators of cell shape, growth, motility
XX and migration. They can also be used to identify proteins that are
XX involved in signal transduction pathways also involving UNC-53, and to
XX identify compounds that alter attachment of UNC-53 to microtubules. A
XX target gene coupled to a UNC-53 encoding sequence may be used to deliver
XX the target gene to a cellular microtubule or its plus ends. The present
XX sequence represents the amino acid sequence of the second human homologue
XX of UNC-53, designated hs-UNC-53/2.
XX
XX Sequence 2432 AA;
XX
XX Query Match 20.5%; Score 55; DB 21; Length 2432;
XX Best Local Similarity 34.0%; Pred. NO. 90;
XX Matches 16; Conservative 3; Mismatches 16; Indels 12; Gaps 1;
XX
XX QY 6 SSPFKYPHTQEAQKQKRSLSG-----EMPGRHLGSSMSLA 40
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 571 smpgkspapsapskegersrsrgklssgipqkqpdldgrhsssssla 617
XX
XX RESULT 12
XX AAY29863
XX ID AAY29863 standard; Protein; 99 AA.
XX
XX AC AAY29863;
XX
XX 17-NOV-1999 (first entry)
XX
XX Human secreted protein clone gn82_6.
XX
XX Human; secreted protein; biological activity; nutritional; cytokine;
XX cell proliferation; differentiation; immune stimulating; vaccine;
XX haematopoiesis regulation; tissue growth; haemostatic; thrombolytic;
XX anti-inflammatory; tumour inhibition.
XX
XX Homo sapiens.
XX
XX OS
XX WO9946287-A1.

```

```

XX 16-SEP-1999.
XX
XX 11-MAR-1999; 99WO-US05243.
XX
XX 11-MAR-1998; 98US-0077521.
XX 14-MAY-1998; 98US-0079124.
XX 10-MAR-1999; 99US-0266105.
XX
XX (GEMY ) GENETICS INST INC.
XX
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
XX Merberg D, Treacy M, Agostino MJ, Steininger RJ;
XX
XX WPI: 1999-551362/46.
XX N-PSDB; AAZ21095.
XX
XX Polynucleotides encoding secreted human proteins, derived from human
XX fetal brain, human adult blood, human adult bladder, or human adult
XX neural tissue cDNA libraries.
XX
XX Claim 15; Page 103; 118pp; English.
XX
XX AAZ21093 to AAZ21102 encode new human secreted proteins and AAY29861 to
XX AAY29873 represent the secreted proteins encoded by the polynucleotide
XX sequences. AAZ21103 to AAZ21112 represent probes for the secreted
XX proteins. The polynucleotides and proteins are predicted to have
XX biological activities which would make them suitable for treating,
XX preventing or ameliorating medical conditions in humans and animals,
XX although no supporting data is given. Suggested activities include
XX nutritional activity, cytokine and cell proliferation/differentiation
XX activity, immune stimulating (e.g. as vaccines) or suppressing activity;
XX haematopoiesis regulating activity, tissue growth activity,
XX activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, receptor/ligand activity, anti-inflammatory
XX activity, cadherin/tumour invasion suppressor activity, and tumour
XX inhibition activity. The polynucleotides and proteins can also be used
XX as nutritional sources or supplements. Such uses include use as a protein
XX or amino acid supplement, use as a carbon source, use as a nitrogen
XX source and use as a source of carbohydrate. They may also have utility
XX in compositions used for bone, cartilage, tendon, ligament, and/or nerve
XX tissue growth or regeneration, as well as for wound healing and tissue
XX repair and replacement, and in the treatment of burns, incisions and
XX ulcers. The proteins which induce cartilage and/or bone growth in
XX circumstances where bone is not normally formed, have application in
XX the healing of bone fractures and cartilage damage or defects in humans
XX and other animals.
XX
XX Sequence 99 AA;
XX
XX Query Match 20.3%; Score 54.5; DB 20; Length 99;
XX Best Local Similarity 36.0%; Pred. No. 2.2;
XX Matches 18; Conservative 7; Mismatches 20; Indels 5; Gaps 2;
XX
XX QY 1 MFLHISPPFKYPHTQEAQKQKRSLSGEMPGRHLGSSMSLALCLVPLVREG 50
XX :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 16 lfllsrtfvlp--vhavkdsaqvleevkkelgsgvsl---lspveepg 60
XX
XX RESULT 13
XX AAY25713
XX ID AAY25713 standard; Protein; 100 AA.
XX
XX AC AAY25713;
XX
XX 04-OCT-1999 (first entry)
XX
XX Human secreted protein encoded from gene 3.
XX
XX Secreted protein; human; treatment; diagnosis; therapy; cancer; tumour;
XX neurodegenerative disorder; developmental abnormality; blood disorder;
XX fetal deficiency; blood disorder; leukemia; immune system; inflammation;
XX
XX KW

```


FT Peptide /note= "WD40 repeat fragment"
FT 127..161
FT Peptide /note= "WD40 repeat fragment"
FT 172..203
FT Peptide /note= "WD40 repeat fragment"
FT 215..252
FT Peptide /note= "WD40 repeat fragment"
FT 265..298
FT Peptide /note= "WD40 repeat fragment"
FT 433..461
FT Peptide /note= "Leucine rich C terminal fragment"

XX JP08111996-A.
XX 14-MAY-1996.
XX 21-OCT-1994; 94JP-0282743.
XX 21-OCT-1994; 94JP-0282743.
XX (NISB) JAPAN TOBACCO INC.
XX WPI; 1996-283507/29.
XX N-ESDB; AAT30361.
XX A new protein, p57, comprising WD40 repeat region - used for
XX development of anti-cancer and anti-HIV agents
XX Example 2; Page 27-29; 5lpp; Japanese.

XX This sequence represents the bovine p57 protein. p57 contains a leucine
XX rich C-terminal peptide which comprises a leucine residue after each
XX 7 amino acids and a WD40 repeated structure region containing five
XX WD40 regions. p57 forms a dimer. It can be used in the development
XX of an anti-cancer agent and an anti-HIV agent. It can also combine
XX with actin in the control of cell movement.
SQ Sequence 461 AA;

Query Match 20.3%; Score 54.5; DB 17; Length 461;
Best Local Similarity 43.8%; Pred. No. 14;
Matches 14; Conservative 3; Mismatches 8; Indels 7; Gaps 1;

Oy 2 FLHSSPFKYPHTQEAQKQKQSLGEMPCRHL 33
||| | : ||| : ||| |
Db 303 flhysmf-----skesqrgmgypkrgl 327

Search completed: July 30, 2001, 16:51:39
Job time: 214 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 04:11:36 ; Search time 6114.61 Seconds
(without alignments)
2894.010 Million cell updates/sec

Title: US-09-402-713A-3
Perfect score: 1872
Sequence: 1 agaagctggcatcagaaaaa.....caataaagaattacaaga 1872

Scoring table: IDENTITY_NUC
Gapop 10_0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues 20456230
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

- 1: gb_estl1:*
- 2: gb_est2:*
- 3: gb_est3:*
- 4: gb_est4:*
- 5: gb_est5:*
- 6: gb_est6:*
- 7: gb_est7:*
- 8: gb_est8:*
- 9: gb_est9:*
- 10: gb_est10:*
- 11: gb_est11:*
- 12: gb_est12:*
- 13: gb_est13:*
- 14: gb_est14:*
- 15: gb_est15:*
- 16: gb_est16:*
- 17: gb_est17:*
- 18: gb_est18:*
- 19: gb_est19:*
- 20: gb_est20:*
- 21: gb_est21:*
- 22: gb_est22:*
- 23: gb_est23:*
- 24: gb_est24:*
- 25: gb_est25:*
- 26: gb_est26:*
- 27: gb_est27:*
- 28: gb_est28:*
- 29: gb_est29:*
- 30: gb_est30:*
- 31: gb_est31:*
- 32: gb_est32:*
- 33: gb_est33:*
- 34: gb_est34:*
- 35: gb_est35:*
- 36: gb_est36:*
- 37: gb_est37:*
- 38: gb_est38:*
- 39: gb_est39:*
- 40: gb_est40:*
- 41: gb_est41:*
- 42: gb_est42:*
- 43: gb_est43:*
- 44: gb_est44:*
- 45: gb_est45:*
- 46: gb_est46:*
- 47: gb_est47:*

- 44: em_esthum10:*
- 45: em_esthum11:*
- 46: em_esthum12:*
- 47: em_esthum13:*
- 48: em_esthum14:*
- 49: em_esthum15:*
- 50: em_esthum16:*
- 51: em_esthum17:*
- 52: em_esthum18:*
- 53: em_esthum19:*
- 54: em_esthum20:*
- 55: em_esthum21:*
- 56: em_esthum22:*
- 57: em_esthum23:*
- 58: em_esthum24:*
- 59: em_esthum25:*
- 60: em_esthum26:*
- 61: em_esthum27:*
- 62: em_esthum28:*
- 63: em_estin1:*
- 64: em_estin2:*
- 65: em_estin3:*
- 66: em_estin4:*
- 67: em_estin5:*
- 68: em_estom1:*
- 69: em_estom2:*
- 70: em_estov1:*
- 71: em_estov2:*
- 72: em_estpl1:*
- 73: em_estpl2:*
- 74: em_estpl3:*
- 75: em_estpl4:*
- 76: em_estpl5:*
- 77: em_estpl6:*
- 78: em_estpl7:*
- 79: em_estpl8:*
- 80: em_estpl9:*
- 81: em_estpl10:*
- 82: em_estro1:*
- 83: em_estro2:*
- 84: em_estro3:*
- 85: em_estro4:*
- 86: em_estro5:*
- 87: em_estro6:*
- 88: em_estro7:*
- 89: em_estro8:*
- 90: em_estro9:*
- 91: em_estro10:*
- 92: em_estro11:*
- 93: em_estro12:*
- 94: em_estro13:*
- 95: em_estro14:*
- 96: em_estro15:*
- 97: em_estro16:*
- 98: em_estro17:*
- 99: em_estro18:*
- 100: em_estro19:*
- 101: em_estro20:*
- 102: gb_est25:*
- 103: gb_est26:*
- 104: gb_est27:*
- 105: gb_est28:*
- 106: gb_est29:*
- 107: gb_est30:*
- 108: gb_est31:*
- 109: gb_est32:*
- 110: gb_est33:*
- 111: gb_est34:*
- 112: gb_est35:*
- 113: gb_est36:*
- 114: gb_est37:*
- 115: gb_est38:*
- 116: gb_est39:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html
Seq primer: 17
Class: BAC ends.

COMMENT

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
c 1	613.8	27.4	654	227	AQ319247	AQ319247 RPCI11-10
c 2	512.2	27.4	856	21	AI557225	AI557225 PT2.1.15
c 3	399.6	21.3	441	170	BF858890	BF858890 RC5-FT019
c 4	394	21.0	402	9	AA578773	AA578773 nh2a04.s
c 5	362	19.3	394	225	AQ206972	AQ206972 HS-3238.B
c 6	306.2	16.4	332	170	BF858286	BF858286 RC5-FT019
c 7	260.4	13.9	290	147	BF373619	BF373619 MR0-FT017
c 8	226.8	12.1	282	147	BF373581	BF373581 MR0-FT017
c 9	214.4	11.5	226	170	BF858371	BF858371 RC5-FT019
c 10	167	8.9	167	147	BF373406	BF373406 IL2-FT015
c 11	138.2	7.4	451	226	AQ246715	AQ246715 HS-2059.B
c 12	86.8	4.6	657	21	AI557495	AI557495 PT2.1.7_H
c 13	63	3.4	1101	219	CNS00396	AL063921 Drosophil
c 14	56	3.0	1001	219	CNS01400	AL103554 Drosophil
c 15	52.6	2.8	1101	219	CNS0182P	AL108811 Drosophil
c 16	52.2	2.8	1101	219	CNS0100X	AL098379 Drosophil
c 17	51.6	2.8	1101	219	CNS000B1	AL085414 Drosophil
c 18	49.2	2.6	1204	219	CNS016E2	AL106628 Drosophil
c 19	48.2	2.6	1101	219	CNS017KX	AL108171 Drosophil
c 20	47.8	2.6	1101	219	CNS0039G	AL063921 Drosophil
c 21	47.6	2.5	529	225	AQ173559	AQ173559 HS-3202.A
c 22	47.6	2.5	1101	219	CNS0039E	AL063919 Drosophil
c 23	47.6	2.5	1101	219	CNS00EVL	AL098706 Drosophil
c 24	47.6	2.5	1101	219	CNS010EY	AL098595 Drosophil
c 25	47.2	2.5	1201	219	CNS0160R	AL106149 Drosophil
c 26	46.8	2.5	654	223	AQ046642	AQ046642 RPCI11-35
c 27	45.2	2.4	1086	219	CNS00YXK	AL096962 Drosophil
c 28	44.8	2.4	938	219	CNS006TJ	AL065906 Drosophil
c 29	44.8	2.4	943	219	CNS002IM	AL097720 Drosophil
c 30	44.8	2.4	1125	106	AL547503	AL547503 AL547503
c 31	44.4	2.4	926	219	CNS008LM	AL052030 Drosophil
c 32	44.4	2.4	1101	219	CNS017KT	AL108167 Drosophil
c 33	43.6	2.3	1101	219	CNS00L72	AL078714 Drosophil
c 34	43.2	2.3	352	12	AA791841	AA791841 vs65b11.r
c 35	43.2	2.3	656	119	AW691330	AW691330 NF043E12S
c 36	43.2	2.3	997	219	CNS0035E	AL060767 Drosophil
c 37	43.2	2.3	1101	219	CNS017KE	AL108152 Drosophil
c 38	43	2.3	928	219	CNS0076D	AL066873 Drosophil
c 39	43	2.3	930	219	CNS00D23	AL060533 Drosophil
c 40	42.8	2.3	821	142	BE960968	BE960968 60164821
c 41	42.8	2.3	1101	219	CNS0039L	AL063926 Drosophil
c 42	42.6	2.3	859	219	CNS004YY	AL055406 Drosophil
c 43	42.6	2.3	1077	222	CNS05AA9	AL328266 Tetraodon
c 44	42.4	2.3	734	219	CNS010MP	AL099163 Drosophil
c 45	42.4	2.3	1092	220	CNS020K7	AL175696 Tetraodon

ALIGNMENTS

RESULT 1
AQ319247/c
LOCUS AQ319247 654 bp DNA GSS 06-MAY-1999
DEFINITION RPCI11-108L4.TV RPCI-11 Homo sapiens genomic clone RPCI-11-108L4, DNA sequence.
ACCESSION AQ319247
VERSION AQ319247.1 GI:4052212
KEYWORDS GSS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 654)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P., and Venter,J.C.
TITLE Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL Unpublished (1998)

```

QY 1629 tatacttcattctctatctatcacaaatatacacaagctttttcacagaattc 1683
|||||
Db 56 TATATTCTATCTCTATCTATCAATATCCACAAGCTTTTCAGAGAAATC 2
|||||

RESULT 2
A1557225 856 bp mRNA EST 09-AUG-1999
LOCUS PT2.L15_B05.r.tumor2 Homo sapiens cDNA 3', mRNA sequence.
DEFINITION A1557225
ACCESSION A1557225
VERSION A1557225.1 GI:4489588
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 856)
AUTHORS Huang, G.M., Ng, W.I., Farkas, J., He, L., Liang, H.A., Gordon, D., Yu, J.
and Hood, L.
TITLE Prostate cancer expression profiling by cDNA sequencing analysis
JOURNAL Genomics 59 (2), 178-186 (1999)
MEDLINE 9333982
COMMENT Contact: Guyang Matthew Huang
Leroy Hood
University of Washington
Department of Molecular Biotechnology, Box 357730, University of
Washington, Seattle, WA 98195
Tel: 5106280100
Fax: 5106280108
Email: huangm@yahoo.com.

FEATURES
source
1..856
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="tumor2"
/note="Organ: Prostate; Vector: pBluescript; Directional
cDNA library was constructed using Lambda ZP II kit
(Stratagene). mRNA was extracted from a frozen prostate
tumor tissue (Mayo Clinics)."
```

BASE COUNT 237 a 191 c 180 g 214 t 34 others

Query Match 27.4%; Score 512.2; DB 21; Length 856;
 Best Local Similarity 96.3%; Pred. No. 4.6e-125;
 Matches 570; Conservative 4; Mismatches 12; Indels 6; Gaps 5;

```

QY 1165 actaagctcttattccctccctctgttgattttttccagataaaagttaaaatgct 1224
|||||
Db 18 ACGTAGCCTTTAATCCCTCCCTTTCTGTTGATTTTTTCCAGTATAAAGTTAAATGCT 77
|||||

QY 1225 tagccttgactgagctgtatag-cacagcctctcccatccctccagccttatctg 1283
|||||
Db 78 TAGCCTTGACTGAGGCTGTATACGCCACAGCCTCTCCCTCCCTCCAGCCTTATCTG 137
|||||

QY 1284 tcatcaccatcaacccctcccatnysacctaaacaaataactgtgtaattccttgaaac 1343
|||||
Db 138 TCATCACCATCAACCCCTCCCATG-CACCTAACAAATCTACTGTGTAATCTCTGAAC 196
|||||

QY 1344 atgtcaggnacatacatttctctgtcgtgagaagctctctgtctcttaantctag 1403
|||||
Db 197 ATGTCAAG-CATACATATTCTCTGCTGAGAGCTCTCTCTCTCTTAAATCTAG 255
|||||

QY 1404 aatgatgtaaaagtttgaataagttgactattcttactcatgcaaaagaaggacacatat 1463
|||||
Db 256 AATGATGTAAGGTTTGAATAAGTTGACATATCTTACTTCATGCAAAAGAGGGACACATAT 315
|||||

QY 1464 gagattcatcatcacatgacagcagcaaaataactaaagtgtaattgtattataagagtta 1523
|||||
Db 316 GAGATTCAATCATCATGACAGACGAAATCTAAAGTGTAAATTTGATTATTAAGAGTTTA 375
|||||

QY 1524 gataaataatgaaatcycaagakccacagaggaatgtttatggggcacgctttgtaagcc 1583
|||||
```

```

Db 376 GATAAATATGAAATGCAAGGCCACAGAGGAATGTTTATGGGCAGCTTTGTAAGCC 435
|||||
QY 1584 tgggatgtgaagmaaagcagaggaaacctcatagtatcttatataataatcttctc 1643
|||||
Db 436 TGGGATGTGAAGCAAAAGCGGAGGAACTCATAGTATCTTATATAATATATCTTCTATTNCTC 495
|||||

QY 1644 tatctctacaaatatacacaagcttttcacagaattcatgcaatcccccaaa 1703
|||||
Db 496 NATCTCTATCAATATCCACAAGCTTTTCACAGAAATTCACGTCAGTGCACAAATCCCAAA 555
|||||

QY 1704 ggtaacctttat-ccatttcattggtgagtcgctttt-agaatttttggcaaa 1752
|||||
Db 556 GGAACCTTTTATCCCATTTTCATGGTGAGTCGCGCTTNAAGAAATTTGGNAAA 607
|||||
```

RESULT 3
 BF858890/c
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

BF858890
 RC5-FT0194-071200-023-G11 FT0194 Homo sapiens cDNA, mRNA sequence.
 BF858890
 BF858890.1 GI:12246634
 EST.
 human.
 Homo sapiens

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 COMMENT

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br

FEATURES
 source

1..441
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="FT0194"
 /dev_stage="Adult"
 /note="Organ: prostate_tumor; Vector: puc18; Site:1: Sma1;
 Site:2: Sma1; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

BASE COUNT 142 a 67 c 97 g 135 t
 ORIGIN

Query Match 21.3%; Score 399.6; DB 170; Length 441;
 Best Local Similarity 97.7%; Pred. No. 3.2e-95;
 Matches 432; Conservative 3; Mismatches 4; Indels 3; Gaps 3;

```

Qy 1128 aatttaattacattttttttccagtgcaagtgactgaagtccttattatccctccct 1187
|||||
Db 441 AATTTAATACATATTTGTTTCCAGTGCAGAGATGACTAGTCTCTTATCCCTCCCT 382
|||||
Qy 1188 ttgtttgattttttccagataaagttaaatccttagccttgactgaggtgtata 1247
|||||
Db 381 TTGTTTGATTTTTTTTCCAGTATAAGTTAAATGCTTAGCTTGTAGCTTGTAGCTGTATA 322
|||||
Qy 1248 cag-cacagcctcccatccctccagccttattgtcattcaccatcaaccctcccat 1306
|||||
Db 321 CAGCCACAGCCTCTCCCATCCCTCCAGCTTATCTGTATCACCATCAACCCCTCCCAT 262
|||||
Qy 1307 nysacttaaaacaaacttaacttgtaattccttgtaacatgtcagncatcatcttct 1366
|||||
Db 261 G-CACCTAAACAAATCTAATCTGTAATCTTGAACATGTCAGG-CATACATTAATCTCT 204
|||||
Qy 1367 ttgctgagagctctctctctctctctcttaantctagaatgatgttaaagttttgataag 1426
|||||
Db 203 TCTGCTGAGAGCTCTCTCTCTCTCTTAATCTAGAATGATGTAAAGTTTGAATAG 144
|||||
Qy 1427 ttgactatcttacttcagcaagaagggacacatagagattcattcaccatcagagaca 1486
|||||
Db 143 TTGACTATCTTACTTCATGCACAAAGGACACATATGAGATTCATCATCATGAGACA 84
|||||
Qy 1487 gcaataactaaaagttaattgatttaaggttttagataaaatatgaaatgcaagak 1546
|||||
Db 83 GCATAATACCAAGTGTAATTTGATTAATAGAGTTTAGATAAATATATGAATGCAAGAG 24
|||||
Qy 1547 ccacagaggggaattgttggtgg 1568
|||||
Db 23 CCACAGAGGGAATGTTTATGGG 2
|||||

RESULT 4
AA578773 402 bp mRNA EST 12-SEP-1997
LOCUS nh24a04.s1 NCI_CGAP_Prl Homo sapiens cDNA clone IMAGE:953262, mRNA
sequence.
ACCESSION AA578773
VERSION AA578773.1 GI:2356957
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 402)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D.
, Michael Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 565 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 373.
Location/Qualifiers
1. 402
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:953262"
/clone.lib="NCI_CGAP_Prl"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"

```

```

/note="Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from 5,000-10
,000 microdissected, histologically normal prostate
epithelial cells. Double-stranded cDNA was ligated to
EcoRI adaptors, 5 cycles of PCR applied to the cDNA with
an adaptor-specific primer, and the resulting PCR product
subcloned into pAMP10 by the uDG-cloning method (Life
Technologies). Average insert size is 600 bp. NOTE: Not
directionally cloned. This library was constructed by
David Krizman."
BASE COUNT 128 a 86 c 84 g 104 t
ORIGIN

Query Match 21.0%; Score 394; DB 9; Length 402;
Best Local Similarity 98.8%; Pred. No. 9.4e-94;
Matches 397; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 395 ggatggggcagaaacagatcctgttggtgatatatttgacgggattacagattga 454
|||||
Db 1 GGTCCACAGGAAACAGATCCTGTGTGGATATATTTGAACGGGATTACAGATTGA 60
|||||
Qy 455 aatgaagtcaaaaagtgaagcattaccgaatgagagaaacacagagaaaaatccttgatgg 514
|||||
Db 61 AATGAAGTCACAAAGTGAAGCATTACCAATGAGAGGAAACACAGCAAGAAATCTTGTATGG 120
|||||
Qy 515 ctccacagacatgcaacaaacaaatggaatactgtgatgacatgagggagccaaagctg 574
|||||
Db 121 CTTCAAGACATGCAACAAACAAATGAATGAACTGTGTATGATGATGAGCAGCCCAAGCTG 180
|||||
Qy 575 gggagagataaccacggggcagaggttcagattctggccctgctgctaaactgctg 634
|||||
Db 181 GGGAGGAGATAACACGGGGCAGGGTCAGGATTTCTGCCCTCTGCTTAACCTGTGGC 240
|||||
Qy 635 ttcaataccaaatcatcttcaatatttcaacctcaaaacaaagctgtgttaatactgat 694
|||||
Db 241 TTCAATACCAAAATCATTTTCATATTTCTAACCTCAAAACAAAGCTGTGTGAATATCTGAT 300
|||||
Qy 695 ctctcaggttctctctggggcccaacattctccatatatccagccacactcattttaata 754
|||||
Db 301 CTCTACGGTTCCTCTGGGGCCCAACATTTCTCATATATATCCAGCCACACTCATTTTAAATA 360
|||||
Qy 755 tttagtccagatctgtactgtgacctttctacactgtaga 796
|||||
Db 361 TTTAGTCTCCAGATCTGTACTGTGACCTTCTACACTGTAGA 402
|||||

RESULT 5
AA0206972/3 394 bp DNA GSS 17-SEP-1998
LOCUS HS_3238_B1_G11_MR CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate=3238 Col=21 Row=N, DNA sequence.
ACCESSION AA0206972
VERSION AA0206972.1 GI:3617542
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 394)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618

```


AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.K., Reis,L.F., de Souza,S.J. and Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

COMMENT Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?ti=MR0st2-MR0-FT0175-310800-106-h09st3-2000-08-31st4-1) Seq primer: puc 18 forward High quality sequence stop: 290 Location/Qualifiers 1. .290 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="FT0175" /dev_stage="Adult" /note="Organ: prostate_tumor; Vector: puc18; Site.1: SmaI; Site.2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions." 78 a 73 c 53 g 86 t

BASE COUNT 78 a 73 c 53 g 86 t

ORIGIN

Query Match 13.9%; Score 260.4; DB 147; Length 290; Best Local Similarity 97.8%; Pred. No. 2.7e-58; Mismatches 0; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 575 gggaggagataaacacggggcagagggtcaggattctggccctgctgctaaactgtgcg 634
Db 21 GGGAGGAGATAACACGGGGCAGAGGGTTCAGGATTCTGGCCCTGCTGCTAAACTGTGCG 80
Qy 635 ttcataaccaaatactttcataattcttaacctcaaaacaaagctgttgttaatactgat 694
Db 81 TTCATAACCAATCATTTTCATATTTCTAACCTCAAAACAAAGCTGTGTAATATCTGAT 140
Qy 695 ctctacggttccctcttgggcccaacattctccatatatccagccacacatcttttaata 754
Db 141 CTCTACGGTTCCTCTTGGGGCCCAACATTTCTCATATATCCAGCCACACTCAATTTTAATA 200
Qy 755 ttttaagttccagatctgactgtgacctcttcaacttagaataaactactctatttg 814
Db 201 TTTAGTTCCTCCAGATCTGACTGTGACCTTTCTACACTGTAGAATAACATTACTCAATTTGG 260
Qy 815 ttcaagacccttcgtgtgctgcctaata 844
Db 261 TTCAAGACCCTTCGTGTGGCTAACGACTA 290

RESULT 8
BF373581/c 282 bp mRNA EST 24-NOV-2000
LOCUS BF373581
DEFINITION MR0-FT0175-210800-101-d05 FT0175 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF373581
VERSION BF373581.1 GI:11335606
KEYWORDS EST.

SOURCE ORGANISM human. Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 282)

AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.K., Reis,L.F., de Souza,S.J. and Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

COMMENT Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?ti=MR0st2-MR0-FT0175-210800-101-d05st3-2000-08-21st4-1) Seq primer: puc 18 forward High quality sequence stop: 282. Location/Qualifiers 1. .282 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="FT0175" /dev_stage="Adult" /note="Organ: prostate_tumor; Vector: puc18; Site.1: SmaI; Site.2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions." 85 a 51 c 70 g 76 t

BASE COUNT 85 a 51 c 70 g 76 t

ORIGIN

Query Match 12.1%; Score 226.8; DB 147; Length 282; Best Local Similarity 91.3%; Pred. No. 2.2e-49; Mismatches 252; Conservative 0; Mismatches 22; Indels 2; Gaps 1;

Qy 568 caagctggggaggagataaacacggggcagagggtcaggattctggccctgctgctaaa 627
Db 282 CAAGCTGGGAGGAGATAAACACGGGGCAGAGGGTTCAGGATTCTGGCCCTGCTGCTAAA 223
Qy 628 ctgtcgctcataaccaaatactttcataattcttaacctcaaaacaaagctgttgtaat 687
Db 222 CTGTGCGTTTCATAACCAATCATTTTCATATTTCTAACCTCAAAACAAAGCTGTGTAAT 163
Qy 688 atctgatctctacggttccctcttgggcccaacattctccatatatccagccacacatt 747
Db 162 ATCTGATCTCTACGGTTCCTCTGCGGCCCAACATTTCTCATATATCCAGCCACACTCAT 103
Qy 748 ttttaataatttagttccagatctgactgtgaccttcttccacacttagaataacatta 805
Db 102 TTTAATATTTAGTTCCTCCAGATCTGACTGTGACCTTTCTCTACATCTGTAGAATAACATTA 43
Qy 806 ctcatctgttcaagacccttcgtgtgctgccta 841
Db 42 CTCATTTGTTCAAGACCCTTCACCTTCTGCTAGCGACTA 7

RESULT 9

BF858371/c 226 bp mRNA EST 16-JAN-2001
 LOCUS RCS-FT0193-211100-012-E11 FT0193 Homo sapiens cDNA, mRNA sequence.
 DEFINITION BF858371
 ACCESSION BF858371
 VERSION BF858371.1 GI:12246115
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 226)
 AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
 Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?cl=RC5&t2=RC5-FT0193-
 211100-012-E11&t3=2000-11-21&t4=1)
 Seq primer: puc 18 forward
 High quality sequence stop: 226.
 Location/Qualifiers
 1..226
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="FT0193"
 /dev_stage="Adult"
 /note="Organ: prostate_tumor; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."
 79 a 29 c 59 g 59 t
 BASE COUNT
 ORIGIN

Query Match 11.5%; Score 214.4; DB 170; Length 226;
 Best Local Similarity 99.5%; Pred. No. 4, 1e-46;
 Matches 215; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 622 cctaaactgctgcgttcataaccacaaatcatttcatttatttccacccctcaaaaagagctgt 681
 Db 226 CCTAAACTGCGTTCATACCAAAATCATTTTCATATTTCTAAACCCCTCAAAACAAAGCTGT 167
 QY 682 tgaataatctgactctacggttctctctctggtggcccaacattctccatatatccagccaca 741
 Db 166 TGAATATCTGATCTCTACGGTTCCTCTCTGCGCCCAACATTCCTCATATATCCAGCCACA 107
 QY 742 ctcatcttttaattagttccagatctgtactgtgacctttctacactgtagaataac 801
 Db 106 CTCATTTTAAATATTAGTTCACATCTGACTGTGACCTTTCTACACTGTAGATAAC 47
 QY 802 attactcattttgttcaagacccttcgtgtgtgtg 837
 Db 46 ATTACTCATTTTGTGTTCAAGACCCCTCGTGTTCGAG 11

RESULT 10
 BF373406
 LOCUS IL2-FT0159-070800-120-H01 FT0159 Homo sapiens cDNA, mRNA sequence.
 DEFINITION BF373406
 ACCESSION BF373406
 VERSION BF373406.1 GI:11335431
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 167)
 AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
 Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?cl=IL2&t2=IL2-FT0159-
 070800-120-H01&t3=2000-08-07&t4=1)
 Seq primer: puc 18 forward
 High quality sequence stop: 167.
 Location/Qualifiers
 1..167
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="FT0159"
 /dev_stage="Adult"
 /note="Organ: prostate_tumor; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."
 47 a 38 c 49 g 33 t
 BASE COUNT
 ORIGIN

Query Match 8.9%; Score 167; DB 147; Length 167;
 Best Local Similarity 100.0%; Pred. No. 1, 5e-33;
 Matches 167; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 254 cacaggaagcacaaggaagcagagatccctgggagaaatgccgcgcacatcttg 313
 Db 1 CACAGGAAGCACAAGGAAGCAGACAGATCCCTGGGAGAAATGCCCGCCGCCATCTTG 60
 QY 314 ggtcactgatgagctcgccctctgctgtgcttcccgcttctgagggagagacattagaaaa 373
 Db 61 GGTCTCATGATGAGCTCGCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
 QY 374 tgaattgatgttctctaaaggatggcaggaagacagatcctgtt 420
 Db 121 TCAATTGATGTGTTCTCTTAAAGGATGGCAGGAAACAGATCCTGTT 167
 RESULT 11
 AQ246715
 LOCUS AQ246715 451 bp DNA GSS 06-OCT-1998

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 07:18:22 ; Search time 176.8 seconds
(without alignments)
2004.471 Million cell updates/sec

Title: US-09-402-713A-3

Perfect score: 1872

Sequence: 1 agaagctggcatcagaaaaa.....caataaagaattacaaga 1872

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCrUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	59.8	3.2	7218	1	US-08-232-463-14
2	57.4	3.1	7218	1	US-08-232-463-14
3	47	2.5	10968	2	US-08-680-327-2
4	47	2.5	10968	4	US-09-228-246-1
5	45.4	2.4	5134	2	US-08-310-912A-157
6	45.4	2.4	5134	5	PCT-US95-04589-157
7	45.4	2.4	5475	2	US-08-680-327-1
8	45.4	2.4	5475	4	US-09-228-246-3
9	35.8	1.9	14507	3	US-08-785-150-1
10	35	1.9	4129	2	US-08-370-319C-12
11	35	1.9	4129	4	US-09-224-834-12
12	34.8	1.9	5718	3	US-08-714-918-48
13	34.8	1.9	5718	4	US-09-265-315-48
14	34.8	1.9	5718	4	US-09-265-315-48
15	34.8	1.9	5718	4	US-09-266-417-48
16	34.2	1.8	2679	1	US-07-977-434-11
17	34.2	1.8	2679	1	US-08-458-819-11
18	34.2	1.8	2679	5	PCT-US91-07035-11
19	33.8	1.8	2570	2	US-09-056-075-2
20	33.8	1.8	246240	2	US-08-724-394A-20
21	33.8	1.8	246240	2	US-08-724-394A-21
22	33.8	1.8	246240	2	US-08-724-394A-22
23	33.6	1.8	3084	1	US-08-551-437-3
24	33.6	1.8	3084	3	US-09-004-225-3
25	33.6	1.8	2084	4	US-09-084-346-3
26	33.2	1.8	5173	1	US-08-242-677-1
27	33	1.8	1952	1	US-08-333-358-1

28	33	1.8	1952	1	US-08-463-694-1	Sequence 1, Appli
29	33	1.8	1952	1	US-08-694-501-1	Sequence 1, Appli
c 30	33	1.8	2950	5	PCT-US93-08386-7	Sequence 7, Appli
c 31	33	1.8	2968	5	PCT-US93-08386-1	Sequence 1, Appli
c 32	33	1.8	8920	2	US-08-446-855A-1	Sequence 1, Appli
c 33	33	1.8	8920	4	US-09-150-741-1	Sequence 1, Appli
c 34	32.8	1.8	720	4	US-08-446-935-4	Sequence 4, Appli
c 35	32.8	1.8	2395	4	US-08-446-935-7	Sequence 7, Appli
c 36	32.8	1.8	2405	1	US-08-454-097-30	Sequence 30, Appli
c 37	32.8	1.8	2405	3	US-08-185-359-30	Sequence 30, Appli
c 38	32.8	1.8	5356	4	US-08-446-935-1	Sequence 1, Appli
39	32.6	1.7	3607	1	US-08-647-351B-1	Sequence 1, Appli
40	32.6	1.7	246240	2	US-08-724-394A-20	Sequence 20, Appli
41	32.6	1.7	246240	2	US-08-724-394A-21	Sequence 21, Appli
42	32.6	1.7	246240	2	US-08-724-394A-22	Sequence 22, Appli
c 43	32.4	1.7	713	3	US-08-532-896-26	Sequence 26, Appli
c 44	32	1.7	2570	2	US-09-056-075-2	Sequence 2, Appli
45	32	1.7	3291	1	US-08-021-601-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

```
Query Match 3.2%; Score 59.8; DB 1; Length 7218;
Best Local Similarity 5.1%; Pred. No. 1.3e-07;
Matches 20; Conservative 219; Mismatches 155; Indels 0; Gaps 0;

QY 1013 ttcatctccacttttggccattctcaagacctcaaaatgctcattccattaatatcac 1072
Db 1076 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1135
QY 1073 aggattaaacttttttaacctgaagaattcaatgtttacatgcagctatgggaattt 1132
Db 1136 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1195
QY 1133 aattacataattttgtttccagtcgaagaatgactaagtcctttatccctccctttgtt 1192
Db 1196 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1255
QY 1193 tgattttttccagtataaagttaaaatgcttagctgtactgaggtgtatcacgca 1252
Db 1256 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1315
QY 1253 cagctctcccatccctcagcttatctgtcatcaccatcaacccctcccatnysacc 1312
Db 1316 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1375
QY 1313 taacaaatctaactgttaattccttgaaatgctcagcagcagcagcagcagcagc 1372
Db 1376 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1435
QY 1373 tgagaagctctcctgtctcttaantctagaat 1406
Db 1436 TACCAAAATCTCTATCTCTCTTAACACTTGTCAT 1469
```

```
RESULT 2
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgt-Fls
US-08-232-463-14

Query Match 3.1%; Score 57.4; DB 1; Length 7218;
Best Local Similarity 4.3%; Pred. No. 6.7e-07;
Matches 16; Conservative 211; Mismatches 142; Indels 0; Gaps 0;

QY 257 aggaagacacaaaggaagcacagagatccctgggagaaatgccgcgccatcttgggt 316
Db 1407 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1348
QY 317 catgatgagctgcctgtgctgctgctgtgaggaagacattagaaaatga 376
Db 1347 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1288
QY 377 attgatgttctctaaagatggcgaggaacagatcctgtgttgatatttga 436
Db 1287 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1228
QY 437 acggattacagattgaaatgaagtcaacaaagtgagcattaccatgagagaaacag 496
Db 1227 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1168
QY 497 acgagaaaaatcttgatgcttcacaaagatgcacacacacacacacacacacacac 556
Db 1167 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1108
QY 557 catgagggcagcaagctgggggagagataacacacacacacacacacacacacacac 616
Db 1107 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1048
QY 617 tgctgccta 625
Db 1047 TGCAGCCAA 1039

RESULT 3
US-08-680-327-2/C
; Sequence 2, Application US/08680327
; Patent No. 5859321
; GENERAL INFORMATION:
; APPLICANT: Staskawicz, Brian S., Oldroyd, Giles Edward,
; APPLICANT: Salmeron, John M., Rommens, Caius
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PLANT
; TITLE OF INVENTION: PATHOGEN RESISTANCE
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klarquist
; STREET: One World Trade Center
; STREET: 121 S.W. Salmon Street
; STREET: Suite 1600
; CITY: Portland
; STATE: Oregon
; COUNTRY: United States of America
; ZIP: 97204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disk, 3-1/2 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/680,327
; FILING DATE: July 11, 1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
```



```

; FILING DATE: April 13, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lech, Karen F.
; REGISTRATION NUMBER: 35,238
; REFERENCE/DOCKET NUMBER: 00786/254001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 100254
; INFORMATION FOR SEQ ID NO: 157:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5134 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-310-912A-157

Query Match 2.4%; Score 45.4; DB 2; Length 5134;
Best Local Similarity 49.7%; Pred. No. 0.0021;
Matches 153; Conservative 4; Mismatches 148; Indels 3; Gaps 2;

QY 1303 coatnysacctaaacaaatctaaattgtaattccttgaaatgtcaggnccatcatrtt 1362
||||: || || || || || || || || || || || || || || || || || || || ||
Db 806 CCATTGTCTTAGGGAAGCTATAAATCTTGAAGAACTCTGGTAC-TTTCAGACAAATAATCTT 748

QY 1363 tctctgctgagaaagctcttctcttcttaantctagaatgatgaaagttttgaa 1422
||||: || || || || || || || || || || || || || || || || || || || ||
Db 747 TACCTCTGCCTCAACAAACTTAAGCAATTCCTTGACATCAGAGAGATAGTGTGTAATA 688

QY 1423 taagtgtactattcttcacgaaggaagacacatgatgattcatcatcacatga 1482
||||: || || || || || || || || || || || || || || || || || || || ||
Db 687 CCAAGCAGGAATATACCCCTTGACAGAGAAACCAAGTCTAGTTTATCTTGAACACG 628

QY 1483 gacagcaataactaaagtgttaattgattataagagtttagataataatgaaatgca 1542
||||: || || || || || || || || || || || || || || || || || || || ||
Db 627 TTCAGCGAGATCTGTAAGTCTATATGTCATATTCGGTACTGTACAAATATCTACAAGTAA 568

QY 1543 agakccacagaggaatgtttatggggcagctttgtaagcctggagtgaaagaaagc 1602
||||: || || || || || || || || || || || || || || || || || || || ||
Db 567 AGAGCCAAATGCAATATTTTGTCTTAACTACTG--AGATTGGTTAAGAGATAAGAC 510

QY 1603 aggggaacc 1610
| |||||
Db 509 ACTGAACC 502

RESULT 6
PCT-US95-04589-157/c
; Sequence 157, Application PC/TUS9504589
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Staskawicz, Brian J.
; APPLICANT: Brent, Andrew F.
; APPLICANT: Dahlbeck, Douglas
; APPLICANT: Katagiri, Fumiki
; APPLICANT: Kunkel, Barbara N.
; APPLICANT: Mindrinos, Michael N.
; APPLICANT: Yu, Guo-Liang
; TITLE OF INVENTION: RPS2 GENE AND USES THEREOF
; NUMBER OF SEQUENCES: 201
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street Suite 3100
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2904
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04589
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/227,360
; FILING DATE: 13-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/230001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 100254
; INFORMATION FOR SEQ ID NO: 157:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5134 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
PCT-US95-04589-157

Query Match 2.4%; Score 45.4; DB 5; Length 5134;
Best Local Similarity 49.7%; Pred. No. 0.0021;
Matches 153; Conservative 4; Mismatches 148; Indels 3; Gaps 2;

QY 1303 coatnysacctaaacaaatctaaattgtaattccttgaaatgtcaggnccatcatrtt 1362
||||: || || || || || || || || || || || || || || || || || || || ||
Db 806 CCATTGTCTTAGGGAAGCTATAAATCTTGAAGAACTCTGGTAC-TTTCAGACAAATAATCTT 748

QY 1363 tctctgctgagaaagctcttctcttcttaantctagaatgatgaaagttttgaa 1422
||||: || || || || || || || || || || || || || || || || || || || ||
Db 747 TACCTCTGCCTCAACAAACTTAAGCAATTCCTTGACATCAGAGAGATAGTGTGTAATA 688

QY 1423 taagtgtactattcttcacgaaggaagacacatgatgattcatcatcacatga 1482
||||: || || || || || || || || || || || || || || || || || || || ||
Db 687 CCAAGCAGGAATATACCCCTTGACAGAGAAACCAAGTCTAGTTTATCTTGAACACG 628

QY 1483 gacagcaataactaaagtgttaattgattataagagtttagataataatgaaatgca 1542
||||: || || || || || || || || || || || || || || || || || || || ||
Db 627 TTCAGCGAGATCTGTAAGTCTATATGTCATATTCGGTACTGTACAAATATCTACAAGTAA 568

QY 1543 agakccacagaggaatgtttatggggcagctttgtaagcctggagtgaaagaaagc 1602
||||: || || || || || || || || || || || || || || || || || || || ||
Db 567 AGAGCCAAATGCAATATTTTGTCTTAACTACTG--AGATTGGTTAAGAGATAAGAC 510

QY 1603 aggggaacc 1610
| |||||
Db 509 ACTGAACC 502

RESULT 7
US-08-680-327-1/c
; Sequence 1, Application US/08680327
; Patent No. 5859321
; GENERAL INFORMATION:
; APPLICANT: Staskawicz, Brian S., Oldroyd, Giles Edward,
; APPLICANT: Salmeron, John M., Rommens, Caius
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PLANT
; TITLE OF INVENTION: PATHOGEN RESISTANCE
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klarquist Sparkman Campbell Leigh &
; ADDRESS: Whinston
; STREET: One World Trade Center
; STREET: 121 S.W. Salmon Street
; STREET: Suite 1600
; CITY: Portland
; STATE: Oregon

```


;; FILING DATE: 11-JAN-96
;; CLASSIFICATION: US/08/586,509
;; PRIOR APPLICATION NUMBER: US/08/586,509
;; FILING DATE: 11-JAN-96
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Perkins, Patricia Anne
;; REGISTRATION NUMBER: 34,693
;; REFERENCE/DOCKET NUMBER: 2841
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (206)587-0430
;; TELEFAX: (206)233-0644
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 14507 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: No. 6027915 Relevant
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; ORIGINAL SOURCE:
;; ORGANISM: Chinese hamster
;; IMMEDIATE SOURCE:
;; CLONE: 2A5-3 lambda CHO sequence
US-08-785-150-1

Query Match 1.9%; Score 35.8; DB 3; Length 14507;
Best Local Similarity 46.9%; Pred. No. 2.9;
Matches 112; Conservative 0; Mismatches 127; Indels 0; Gaps 0;
QY 870 gtgttctgcccagggtatctgtgaacaggtctggaagcatctcaagatctttccagggt 929
Db 10959 GGGTCTTCCACAGTGGATGATGCTCATCAAGGGAATTTATTCACCCGTGGTGTAGTG 10900
QY 930 tatactactagcacagcatgatctaccaggtgagtggaattatcttaatacaacatcatcct 989
Db 10899 GCTACAATAGCCGGCTAGAGGCGACAGACGCTTCTTAATCAGGTCCTGTGTGTAC 10840
QY 990 cagtgtcttgccatactgaattcatcttccacatttggccatttccagacacctca 1049
Db 10839 GAGATGACTAGCGGCAAGTATACAGCGCAACGCTTTGAGAAATTTAAAGTGTCTCA 10780
QY 1050 aaatgtcattccattatcaacagattaaacttttttttaacccgtggaagaattcaa 1108
Db 10779 CAACGAGTTCAGTAGTAGACAGACAGAAACATTTATTTTATTTGGAAATAATATA 10721

RESULT 10
US-08-370-319C-12/c
; Sequence 12, Application US/08370319C
; Patent No. 5856091
; GENERAL INFORMATION:
; APPLICANT: Brichard, Vincent; Van Pel, Aline;
; APPLICANT: Traversari, Catia; W lfel, Thomas; Coullie, Pierre;
; APPLICANT: Boon-Falleur, Thierry; De plaen, Etienne
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID SEQUENCE CODING FOR A
; TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PRECURSOR PROCESSED TO AT LEAST ONE TU
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/370,319C
;; FILING DATE: 10-JANUARY-1995
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/272,351
;; FILING DATE: 8-JULY-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/032,978
;; FILING DATE: 18-MAR-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Hanson, No. 5856091man D.
;; REGISTRATION NUMBER: 30,946
;; REFERENCE/DOCKET NUMBER: LUD 5377.1
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 688-9200
;; TELEFAX: (212) 838-3884
;; INFORMATION FOR SEQ ID NO: 12:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 4129 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; FEATURE:
;; OTHER INFORMATION: The sequence is preceded by an
;; OTHER INFORMATION: unsequenced portion of from 4.7 to 5.3
;; OTHER INFORMATION: kilobases
US-08-370-319C-12

Query Match 1.9%; Score 35; DB 2; Length 4129;
Best Local Similarity 48.7%; Pred. No. 2.4;
Matches 95; Conservative 0; Mismatches 100; Indels 0; Gaps 0;
QY 1025 ttttggccattctcaagacacctcaaaatgtcattccattatatacacaggattacttt 1084
Db 2703 TCTTTTCTACATACTAATAATTTTGAATATGATATGATATTAATAATACATGCTTATTTC 2644
QY 1085 ttttttaacctgggaattcaatgtttacatgcagctatgggaatttaattacatat 1144
Db 2643 AACTCTGTGTTTAAAGTGTTCATAGCACCAGTACTGAGTGGCTAAGCCATCAGACAG 2584
QY 1145 tgttttccagtgcaagatgactaagtccttccctccctcccttggttgttttttttc 1204
Db 2583 CACAGCTCTATACAATAATAATAGTACTTTTCCCTCCCTCTGTATATATCTGCATC 2524
QY 1205 cagtataaagttaaa 1219
Db 2523 CATTAAGGAGGCAAA 2509

RESULT 11
US-09-224-834-12/c
; Sequence 12, Application US/09224834
; Patent No. 6201111
; GENERAL INFORMATION:
; APPLICANT: Brichard, Vincent; Van Pel, Aline;
; APPLICANT: Traversari, Catia; W lfel, Thomas; Coullie, Pierre;
; APPLICANT: Boon-Falleur, Thierry; De plaen, Etienne
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID SEQUENCE CODING FOR A
; TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PRECURSOR PROCESSED TO AT LEAST ONE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM PS/2
; SOFTWARE: IBM PS/2

APPLICANT: Martin, Patrick K.
APPLICANT: Schmid, Molly B.
APPLICANT: Sun, Dongxu
TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
TITLE OF INVENTION: TARGET GENES
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/266,417
FILING DATE: March 9, 1999
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/714,918
FILING DATE: September 13, 1996
APPLICATION NUMBER: 60/009,102
FILING DATE: December 22, 1995
APPLICATION NUMBER: 60/003,798
FILING DATE: September 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 240/248
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 5718 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-266-417-48

Query Match 1.9%; Score 34.8; DB 4; Length 5718;
Best Local Similarity 48.1%; Pred. No. 3.4;
Matches 99; Conservative 0; Mismatches 107; Indels 0; Gaps 0;
QY 1033 ccatttcagaacctcaaatgtccattccatttaataatcacaggattaaactttttttta 1092
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 283 CCATCAGTAATAACTGCAACTGATTTCCTTTAAATGTGTAATCATATACTTTTCTTTTA 342
QY 1093 acctggaagaattcaattgatacagctatgggaatttaattacattttgttttcc 1152
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 343 TCTTCATAAATATCTTTACAGGTTACGCCAGCCAGGTGAGTATGCTAAACTTAATCC 402
QY 1153 agtgcagaagtactaaagtccctttccctccctctgtttgatttttttccagataa 1212
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 403 TCTTTATTAGTAACCTTTTACATTGTTTAACTTCTAATTTACCTTGATACGTTTGTGC 462
QY 1213 agttaaatgcttagcctgtactga 1238
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 463 ATTCCAATGCTTCATCTCTTAATGA 488

Search completed: July 31, 2001, 07:20:07
Job time: 17569 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 07:25:26 ; Search time 478.87 Seconds
(without alignments)
2454.595 Million cell updates/sec

Title: US-09-402-713A-3

Perfect score: 1872

Sequence: 1 agaagctgcatcagaaaaa.....caataagaattacaaga 1872

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_0601.*
1: /SID88/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1867	99.7	1872	19 AAV62428	Prostate cancer an
2	1772.4	94.7	2037	19 AAV62427	Prostate cancer an
3	1772.4	94.7	3582	19 AAV62430	Prostate cancer an
4	1739.2	92.9	3112	21 AAA06687	Human immunogenic
c 5	1734	92.6	2426	21 AAA06689	Human immunogenic
c 6	1715	91.6	2229	21 AAA06688	Human immunogenic
7	820	43.8	820	19 AAV62429	Prostate cancer an
c 8	812	43.4	812	21 AAA06690	Human immunogenic
9	524	28.0	597	20 AAX37486	Human secreted pro
10	513.2	27.4	718	21 AAA06545	Human immunogenic
11	414.2	22.1	437	21 AAC06768	Human secreted pro

c 12	330	17.6	359	20	AA233445	Human prostate can
c 13	288.4	15.4	301	21	AA06520	Human immunogenic
c 14	136	7.3	936	22	AAF58252	Oligonucleotide D1
c 15	136	7.3	936	22	AAF58254	Oligonucleotide D1
c 16	136	7.3	936	22	AAF58257	Oligonucleotide D1
c 17	136	7.3	936	22	AAF58259	Oligonucleotide D2
c 18	136	7.3	936	22	AAF58262	Oligonucleotide D2
c 19	136	7.3	936	22	AAF58255	Oligonucleotide D1
c 20	134.4	7.2	936	22	AAF58252	Oligonucleotide D1
c 21	134.4	7.2	936	22	AAF58254	Oligonucleotide D1
c 22	134.4	7.2	936	22	AAF58257	Oligonucleotide D1
c 23	134.4	7.2	936	22	AAF58259	Oligonucleotide D2
c 24	134.4	7.2	936	22	AAF58262	Oligonucleotide D2
c 25	134.4	7.2	936	22	AAF58255	Oligonucleotide D1
c 26	100	5.3	123	19	AAV33791	Prostate cancer an
c 27	50.8	2.7	244	22	AAF58238	Oligonucleotide D1
c 28	47.6	2.5	244	22	AAF58238	Oligonucleotide D1
c 29	47	2.5	10968	19	AAV17789	Tomato Prf genomic
c 30	45.4	2.4	5475	19	AAV17777	Tomato Prf cDNA.
c 31	40	2.1	796	20	AAV37484	Human secreted pro
c 32	39	2.1	2206	21	AA12390	Human interferon-b
c 33	39	2.1	10325	20	AAV76516	Plasmid containing
c 34	37.6	2.0	1297	13	AAQ21645	3' coding sequence
c 35	37.6	2.0	1770	14	AAQ37684	P.falciiparum antig
c 36	37.6	2.0	1770	14	AAQ37124	P.falciiparum antig
c 37	37.4	2.0	1255	22	AAF72762	Human prostate can
c 38	37.4	2.0	1325	21	AAF15730	Human prostate can
c 39	37.4	2.0	1325	22	AAF72787	Human prostate can
c 40	37.2	2.0	1751	21	AAV58122	Human secreted pro
c 41	36.8	2.0	5804	18	AAV74560	Staphylococcus aur
c 42	36.4	1.9	2643	15	AAQ66096	S. cerevisiae Plc1
c 43	36.2	1.9	1451	21	AAF21739	Human breast and o
c 44	36.2	1.9	1517	21	AAC34792	Arabidopsis thalia
c 45	36.2	1.9	1630	20	AAV13310	Enterococcus faeca

ALIGNMENTS

RESULT 1

AAV62428

ID : AAV62428 standard; cDNA; 1872 BP.

XX

AC AAV62428;

XX

DT 30-DEC-1998 (first entry)

XX

DE Prostate cancer antigen (PCA3) cDNA splice variant 2.

XX

KW Prostate cancer antigen cDNA splice variant 2; PCA3; prostatic cancer;

PC; ds.

XX

OS Homo sapiens.

XX

PN WO9845420-A1.

XX

PD 15-OCT-1998.

XX

PF 09-APR-1998; 98WO-CA00345.

XX

PR 10-APR-1997; 97US-0041836.

XX

PA (DIAG-) DIAGNOCURE INC.

XX

PI Bussemakers MJG;

XX

DR WPI; 1998-568347/48.

XX

PT New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,

prevention and treatment of prostatic cancer

XX

PS Claim 4; Pages 76-77; 111pp; English.

XX

CC The present sequence represents the prostate cancer antigen (PCA3)
CC cDNA splice variant 2 sequence comprising of exons 1, 3, 4a and
CC 4b of the PCA3 gene. The PCA3 cDNA splice variant 2 sequence,
CC isolated from a human primary prostatic tumour tissue cDNA library,
CC was found in approximately 65% of the cDNA clones isolated. The
CC invention claims for PCA3 cDNA variants and the proteins they encode.
CC The invention also claims for antibodies against PCA3 protein. In
CC antibodies are also claimed to be useful for detecting PCA3 protein in
CC immunoassay tests, for diagnosing, assessing and prognosing of
CC prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin
CC or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed
CC to be useful for treating PC, while determining elevated levels of
CC PCA3 (as RNA or protein) is useful for detecting a predisposition
CC to development of PC, e.g. in prenatal tests. Detecting PCA3 protein
CC allows differentiation between malignant and benign prostatic disease,
CC and the level of PCA3 expression allows correlation with the grade of
CC tumour. PCA3 protein and its fragments are also claimed to be useful
CC in vaccines for preventing PC; in drug screens for identifying
CC specific (antagonists (potentially useful therapeutically) and for
CC studying protein-DNA interactions.
XX
SQ Sequence 1872 BP; 567 A; 389 C; 369 G; 539 T; 8 other;

Query Match 99.7%; Score 1867; DB 19; Length 1872;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1872; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 agaagctggcatcagaaaaacagagggagattgtgctgcgacccgagggagaccag 60
DB 1 agaagctggcatcagaaaaacagagggagattgtgctgcgacccgagggagaccag 60
QY 61 gaagatctgcattgtgggaagacacctgatgacagaggtgagaaataagaagacctgt 120
DB 61 gaagatctgcattgtgggaagacacctgatgacagaggtgagaaataagaagacctgt 120
QY 121 gactttaccatctgagccacacatctgctgaaatggagataattaacatcactagaaac 180
DB 121 gactttaccatctgagccacacatctgctgaaatggagataattaacatcactagaaac 180
QY 181 agcaagatgacaaataaattgctaaagtagtgacatgtttttcacatttccagcccttt 240
DB 181 agcaagatgacaaataaattgctaaagtagtgacatgtttttcacatttccagcccttt 240
QY 241 aaatattccacacacaggaagacaaaaaggaagcacagagatccctgggagaaatgcc 300
DB 241 aaatattccacacacaggaagacaaaaaggaagcacagagatccctgggagaaatgcc 300
QY 301 ggcgcctatttgggtcatcgatgagcctcgccctgtgctgttcccgcttggaggaa 360
DB 301 ggcgcctatttgggtcatcgatgagcctcgccctgtgctgttcccgcttggaggaa 360
QY 361 ggacattagaaaatgaattgatgtgttccttaaaagtgaggcaggaacacagatcctgt 420
DB 361 ggacattagaaaatgaattgatgtgttccttaaaagtgaggcaggaacacagatcctgt 420
QY 421 gtggatatttatttgaacgggattacagatttgaatgaagtcaaaaagtggacattacc 480
DB 421 gtggatatttatttgaacgggattacagatttgaatgaagtcaaaaagtggacattacc 480
QY 481 aatgagagaaacacagacagaaaaatttgtatgcttcaacagacatgcaacaaacaaa 540
DB 481 aatgagagaaacacagacagaaaaatttgtatgcttcaacagacatgcaacaaacaaa 540
QY 541 tggaaactgtgtgacatgagcgaccaaagctggggaggagataaccacggggcagagg 600
DB 541 tggaaactgtgtgacatgagcgaccaaagctggggaggagataaccacggggcagagg 600
QY 601 gtcaagatttctgctgtcgttaaaactgtcgttcaataaccaaatacatttcatttc 660
DB 601 gtcaagatttctgctgtcgttaaaactgtcgttcaataaccaaatacatttcatttc 660
QY 661 taacctcaaaacaaagctgttgtaattctgatctctacggttctctctggggccaaca 720

DB 661 taacctcaaaacaaagctgttgtaattctgatctctacggttctctctggggccaaca 720
QY 721 ttctccatatatccacccacactcattttaattatttagttccagatctgtactgtac 780
DB 721 ttctccatatatccacccacactcattttaattatttagttccagatctgtactgtac 780
QY 781 cttctcacactgtagaataacattactctatttggttcaaaagaccctctgtgtgctgcct 840
DB 781 cttctcacactgtagaataacattactctatttggttcaaaagaccctctgtgtgctgcct 840
QY 841 aatatgtactgactgttttttcttaaggagtggttctggcccgagggatctgtgaacaggc 900
DB 841 aatatgtactgactgttttttcttaaggagtggttctggcccgagggatctgtgaacaggc 900
QY 901 tgggaagacatccaagatcttttccaggggttatacttactagcacacagatgatattac 960
DB 901 tgggaagacatccaagatcttttccaggggttatacttactagcacacagatgatattac 960
QY 961 ggagtgaaattatctaatacacaatcctcagtgcttggcccatcactgaaattcatttc 1020
DB 961 ggagtgaaattatctaatacacaatcctcagtgcttggcccatcactgaaattcatttc 1020
QY 1021 ccacttttggccattctcaagacccctcaaaatgtctatccatttaataatcacaggattaa 1080
DB 1021 ccacttttggccattctcaagacccctcaaaatgtctatccatttaataatcacaggattaa 1080
QY 1081 ctttttttttaacctggagaattcaattgattacatgcagctatgggaatttaattacat 1140
DB 1081 ctttttttttaacctggagaattcaattgattacatgcagctatgggaatttaattacat 1140
QY 1141 atttggtttccagtgcaaaagatgactaagtcctttatccctccctttgtttgatttt 1200
DB 1141 atttggtttccagtgcaaaagatgactaagtcctttatccctccctttgtttgatttt 1200
QY 1201 ttccagtgataaagttaaaatgcttagcctgttactgagtggtgtgtacagcacagcctct 1260
DB 1201 ttccagtgataaagttaaaatgcttagcctgttactgagtggtgtgtacagcacagcctct 1260
QY 1261 ccccatccctccagccttactgtcatcaccatcaacccctcccatcctcaacaaacaa 1320
DB 1261 ccccatccctccagccttactgtcatcaccatcaacccctcccatcctcaacaaacaa 1320
QY 1321 atctaacttgtaattccttgaacatgtcaggncatcacatrttctcttgcctgagaagc 1380
DB 1321 atctaacttgtaattccttgaacatgtcaggncatcacatrttctcttgcctgagaagc 1380
QY 1381 tcttctgtcttcttaantctagaatgtgtaaaagtttgaataagttgactatcttact 1440
DB 1381 tcttctgtcttcttaantctagaatgtgtaaaagtttgaataagttgactatcttact 1440
QY 1441 tcatcaagaagaaggacacacatgatgatttcacatcacatgagacagcaaaactactaaag 1500
DB 1441 tcatcaagaagaaggacacacatgatgatttcacatcacatgagacagcaaaactactaaag 1500
QY 1501 tgaatttgattataaagtttagataaaatatgaatgaagkccacacagaggaatg 1560
DB 1501 tgaatttgattataaagtttagataaaatatgaatgaagkccacacagaggaatg 1560
QY 1561 tttatggggcagcttgttaagcctgggatgtgaamaaggcagggaacctcatagtc 1620
DB 1561 tttatggggcagcttgttaagcctgggatgtgaamaaggcagggaacctcatagtc 1620
QY 1621 ttataataatacttcttctctctctatcctatccaaatccaaatccaaagcttttcacaaa 1680
DB 1621 ttataataatacttcttctctctctatcctatccaaatccaaagcttttcacaaa 1680
QY 1681 ttcagcagtgcaaatcccccaaggtaacctttatccatttcatggtgagtcgctttag 1740
DB 1681 ttcagcagtgcaaatcccccaaggtaacctttatccatttcatggtgagtcgctttag 1740
QY 1741 aatttggcaaatcactactggtcacttctcacttatcctcaatttgagatgtgttctgtgtagt 1800

Db 1741 aatttggcaaatcactactgtgctcacttatctcaactttagagatgtgttctgttagt 1800
QY 1801 taattgaagaataagggcactctgtgagccactttaggggtcactctcggcaataaag 1860
Db 1801 ttt 1860
QY 1861 aatttacaaga 1872
Db 1861 aatttacaaga 1872
RESULT 2
AAV62427
ID AAV62427 standard; cDNA; 2037 BP.
XX
AC AAV62427;
XX 30-DEC-1998 (first entry)
XX
XX Prostate cancer antigen (PCA3) cDNA splice variant 1.
XX Prostate cancer antigen cDNA splice variant 1; PCA3; prostatic cancer;
KW PC; ds.
XX
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 379..534
FT /*tag= a
FT /product= "PCA3 protein variant 1"
FT 2019..2024
FT polyA_signal /*tag= b
XX
XX WO9845420-A1.
XX
XX 15-OCT-1998.
XX
XX 09-APR-1998; 98WO-CA00346.
XX
XX 10-APR-1997; 97US-0041836.
XX
XX (DIAG-) DIAGNOCURE INC.
XX
XX Bussemakers MJG;
XX
XX WPI; 1998-568347/48.
XX P-P5DB; AAW79736.
XX
XX New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
XX prevention and treatment of prostatic cancer
XX
XX Claim 3; Fig 2B-2J; 11pp; English.
XX
XX The present sequence represents the prostate cancer antigen (PCA3)
XX cDNA splice variant 1 sequence comprising of exons 1, 2, 3, 4a and
XX 4b of the PCA3 gene. The PCA3 cDNA splice variant 1 sequence,
XX isolated from a human primary prostatic tumour tissue CDNA library,
XX was found in approximately 5% of the CDNA clones isolated. The
XX invention claims for PCA3 cDNA variants and the proteins they encode.
XX The invention also claims for antibodies against PCA3 protein. The
XX antibodies are claimed to be useful for detecting PCA3 protein in
XX immunoassay tests, for diagnosing, assessing and prognosing of
XX prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin
XX or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed
XX to be useful for treating PC, while determining elevated levels of
XX PCA3 (as RNA or protein) is useful for detecting a predisposition
XX to development of PC, e.g. in prenatal tests. Detecting PCA3 protein
XX allows differentiation between malignant and benign prostatic disease,
XX and the level of PCA3 expression allows correlation with the grade of
XX tumour. PCA3 protein and its fragments are also claimed to be useful
XX in vaccines for preventing PC; in drug screens for identifying
XX specific (ant)agonists (potentially useful therapeutically) and for
XX studying protein-DNA interactions.

XX Sequence 2037 BP; 622 A; 426 C; 406 G; 575 T; 8 other;
SQ
Query Match 94.7%; Score 1772.4; DB 19; Length 2037;
Best Local Similarity 99.9%; Pred. NO. 0;
Matches 1778; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 94 cagagtgagaaataagaaagctgctgactttacactctgagggccacacatctcgtgaa 153
Db 259 caggggtgagaaataagaaagctgctgactttacactctgagggccacacatctcgtgaa 318
QY 154 atggagataaattacatcactagaaacagagatgacaataataatgtctaagtgtgac 213
Db 319 atggagataaattacatcactagaaacagagatgacaataataatgtctaagtgtgac 378
QY 214 atgttttgcacatttcagcccccttttaaatatcacacacacaggaagcacaagaagaa 273
Db 379 atgttttgcacatttcagcccccttttaaatatcacacacacaggaagcacaagaagaa 438
QY 274 gcacagagatccctgggagaaatgcccgccctcttggctgcatcgatgagcctcgcc 333
Db 439 gcacagagatccctgggagaaatgcccgccctcttggctgcatcgatgagcctcgcc 498
QY 334 ctgtgctgtgctccctgtgtgaggaagacattagaaatgaattgatgtgttccttaa 393
Db 499 ctgtgctgtgctccctgtgtgaggaagacattagaaatgaattgatgtgttccttaa 558
QY 394 aggtatggcaggaaacacagatcctgtgtggtattattttgaacgggattacagattcg 453
Db 559 aggtatggcaggaaacacagatcctgtgtggtattattttgaacgggattacagattcg 618
QY 454 aaatgaagtcaaaagtgaagcattaccatgagagaaacacagaaataatcttgatg 513
Db 619 aaatgaagtcaaaagtgaagcattaccatgagagaaacacagaaataatcttgatg 578
QY 514 gcttcacaagacatgcaacaaacaaatggaatactgtgtgacatgagcgagccaagct 573
Db 679 gcttcacaagacatgcaacaaacaaatggaatactgtgtgacatgagcgagccaagct 738
QY 574 gggggaggagataaacacagggggcagaggggtcaggattcttggcctgctgctaaactgtgc 633
Db 739 gggggaggagataaacacagggggcagaggggtcaggattcttggcctgctgctaaactgtgc 798
QY 634 gtccataaccacaaatcatttcataatttccaccctcaaaacaaagctgtgttaatactga 693
Db 799 gtccataaccacaaatcatttcataatttccaccctcaaaacaaagctgtgttaatactga 858
QY 694 tctctacggttctcttctggcccaacattctccatatatccagccacactcattttaat 753
Db 859 tctctacggttctcttctggcccaacattctccatatatccagccacactcattttaat 918
QY 754 attagtctccagatgtgtactgtgacctttctaacactgtagaataacattactcatttt 813
Db 919 attagtctccagatgtgtactgtgacctttctaacactgtagaataacattactcatttt 978
QY 814 gtccaaagacccttgggtgtgctgctcctaataatgtagtgactgtttttcctaaggagtgt 873
Db 979 gtccaaagacccttgggtgtgctgctcctaataatgtagtgactgtttttcctaaggagtgt 1038
QY 874 tctgcccaggagatctgtgaacagggctgggaagcatctcaagatcttccagggttata 933
Db 1039 tctgcccaggagatctgtgaacagggctgggaagcatctcaagatcttccagggttata 1098
QY 934 cttaactagcacacagatgatcattacggagtgaaattatctaatcaacatcatcctcagt 993
Db 1099 cttaactagcacacagatgatcattacggagtgaaattatctaatcaacatcatcctcagt 1158
QY 994 gtctttgcccatactgaaattcatttccacacttttggccatttctcaagacactcaaat 1053
Db 1159 gtctttgcccatactgaaattcatttccacacttttggccatttctcaagacactcaaat 1218
QY 1054 gtcatccatataatcacagaggaattaaacttttttttaacctgggaagattcaattgta 1113

```
|||||
Db 1219 gtcattccatataatcacaggattaaacttttttttaacctggaagaattcaaatgtta 1278
QY 1114 catgagctatgggaatttaattacataattttgttttccagtgcaaaagatgactaagtc 1173
Db 1279 catgagctatgggaatttaattacataattttgttttccagtgcaaaagatgactaagtc 1338
QY 1174 ttatccctccctctgtgtgtgttttttccagtataaagttaaaatcttagccttgt 1233
Db 1339 ttatccctccctctgtgtgtgttttttccagtataaagttaaaatcttagccttgt 1398
QY 1234 actgaggtgtatcacgacagcctctccctccatccctccagccttatctgttcatcaccat 1293
Db 1399 actgaggtgtatcacgacagcctctccctccatccctccagccttatctgttcatcaccat 1458
QY 1294 caaccctccctatnysacctaaacaaaactaactgttaattccttgaacatgtcaggnc 1353
Db 1459 caaccctccctatnysacctaaacaaaactaactgttaattccttgaacatgtcaggnc 1518
QY 1354 atacatrrtctctgtcgtgagaagctctctctgtctcttaantctagaatgatgtaa 1413
Db 1519 atacatrrtctctgtcgtgagaagctctctctgtctcttaantctagaatgatgtaa 1578
QY 1414 agtttgaataagttgactatcttacttcattgcgaagaaggagacacatatgagattcatc 1473
Db 1579 agtttgaataagttgactatcttacttcattgcgaagaaggagacacatatgagattcatc 1638
QY 1474 atcacatgagacagcaaatactaaaagttaattgtattataagatttagataaaata 1533
Db 1639 atcacatgagacagcaaaactaaaagttaattgtattataagatttagataaaata 1698
QY 1534 tgaattgcaagkacacagaggaattgtttatgggcacgtttgttaagcctgggatgtga 1593
Db 1699 tgaattgcaagkacacagaggaattgtttatgggcacgtttgttaagcctgggatgtga 1758
QY 1594 agmaaggcagggaacctcatagttatcttataataataatacttctctctctctctc 1653
Db 1759 agmaaggcagggaacctcatagttatcttataataataatacttctctctctctc 1818
QY 1654 acaatatcaacaagcttttcagaattcgcagtgcaaatcccaaggttaaccttt 1713
Db 1819 acaatatcaacaagcttttcagaattcgcagtgcaaatcccaaggttaaccttt 1878
QY 1714 atccatttcattgtagtgcgttttagaattttggcaaatcactactggtcacttatctca 1773
Db 1879 atccatttcattgtagtgcgttttagaattttggcaaatcactactggtcacttatctca 1938
QY 1774 actttgagatgtgttgccttgcctgtagtttaattgaaagaaatagggcactcttctgagcca 1833
Db 1939 actttgagatgtgttgccttgcctgtagtttaattgaaagaaatagggcactcttctgagcca 1998
QY 1834 ctttaggggttcactcctcgtgcaataaagaattttacaaaga 1872
Db 1999 ctttaggggttcactcctcgtgcaataaagaattttacaaaga 2037
```

RESULT 3

AAV62430
ID AAV62430 standard; cDNA; 3582 BP.

XX AAV62430;

XX 30-DEC-1998 (first entry)

XX Prostate cancer antigen (PCA3) wild-type cDNA.

XX Prostate cancer antigen cDNA; PCA3; prostatic cancer;
KW PC; ds.

XX Homo sapiens.

XX Key Location/Qualifiers
FT CDS 401..556

FT /*tag= a
FT /product= "PCA3 protein"
FT 983..987
FT /*tag= b
FT polyA_signal
FT 2041..2046
FT /*tag= c
FT polyA_signal
FT 2597..2602
FT /*tag= d
FT polyA_signal
FT 3494..3496
FT /*tag= e

XX WO9845420-A1.

PN 15-OCT-1998.

XX 09-APR-1998; 98WO-CA00346.

XX 10-APR-1997; 97US-0041836.

XX (DIAG-) DIAGNOCURE INC.

XX Bussemakers MJG;

XX WPI; 1998-568347/48.
DR P-PSDB; AAW79738.

XX New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
prevention and treatment of prostatic cancer

XX Claim 3; Fig 5B-5F; lllpp; English.

XX The present sequence represents the prostate cancer antigen (PCA3)
wild-type cDNA sequence comprising of exons 1, 2, 3, 4a-4g of the
PCA3 gene. The invention claims for PCA3 cDNA variants and the
proteins they encode. The invention also claims for antibodies
against PCA3 protein. The antibodies are claimed to be useful for
detecting PCA3 protein in immunoassay tests, for diagnosing, assessing
and prognosing of prostatic cancer (PC). Antibodies, optionally
coupled to a cytotoxin or radioisotope, and nucleic acids antisense
to PCA3 cDNA are claimed to be useful for treating PC, while determining
elevated levels of PCA3 (as RNA or protein) is useful for detecting a
predisposition to development of PC, e.g. in prenatal tests. Detecting
PCA3 protein allows differentiation between malignant and benign
prostatic disease, and the level of PCA3 expression allows correlation
with the grade of tumour. PCA3 protein and its fragments are also
claimed to be useful in vaccines for preventing PC; in drug screens
for identifying specific (ant)agonists (potentially useful
therapeutically) and for studying protein-DNA interactions.

XX Sequence 3582 BP; 1052 A; 788 C; 679 G; 1063 T; 0 other;

Query Match 94.7%; Score 1772.4; DB 19; Length 3582;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1770; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 94 cagaggtgagaaataagaaggctgctgactttaccatctgagggccacacatctgctgaa 153
Db 281 cagggtgagaaataagaaggctgctgactttaccatctgagggccacacatctgctgaa 340
QY 154 atggagataatttaacatcatcactagaacagagatgacaataatgtctaagtagtac 213
Db 341 atggagataatttaacatcatcactagaacagagatgacaataatgtctaagtagtac 400
QY 214 atgtttttgcacatttcagcccttttaataatccacacagagaaacacaaaggaa 273
Db 401 atgtttttgcacatttcagcccttttaataatccacacagagaaacacaaaggaa 460
QY 274 gcacagagatccctgggagaaatgccggccgcctcttgggtcactcgatgagcctcgcc 333
Db 461 gcacagagatccctgggagaaatgccggccgcctcttgggtcactcgatgagcctcgcc 520
QY 334 ctgtgcctggtccccctgtgagggaaggagacattagaaaaatgaattgattgttctcttaa 393

Db 521 cttgtcctgttccgcctgtgagggaaggacattagaaaaatgattgattgttctcttaa 580
Qy 394 aggatggcaggaacacagatcctgttggatattttatttgaaacggattacagatttg 453
Db 581 agga tggcaggaacacaga tccgtgttgatattttatttgaaacggattacagatttg 640
Qy 454 aaatgaagtccaaagttagcattaccatgagagaaacacagacgagaaatcttgatg 513
Db 641 aaatgaagtccaaagttagcattaccatgagagaaacacagacgagaaatcttgatg 700
Qy 514 gcttcacaagacatgcaacaaacaaaaatggaatactgtgtagacatgaggcagcaagct 573
Db 701 gcttcacaagacatgcaacaaacaaaaatggaatactgtgtagacatgaggcagcaagct 760
Qy 574 ggggagagataaacacacgggagagaggtcagattctggccctgtcctcaaacgtgac 633
Db 761 ggggagagataaacacacgggagagaggtcagattctggccctgtcctcaaacgtgac 820
Qy 634 gttcataaccacaaatcatcttcatattctaaacctcaacaaacagctgtgtaatactga 693
Db 821 gttcataaccacaaatcatcttcatattctaaacctcaacaaacagctgtgtaatactga 880
Qy 694 tctctacaggttctctctgggcccacattctccatataatccagccacactcatctttaat 753
Db 881 tctctacaggttctctctgggcccacattctccatataatccagccacactcatctttaat 940
Qy 754 atttagttccacagatctgtactgtgacattctctacactgtagaataacattactcat 813
Db 941 atttagttccacagatctgtactgtgacattctctacactgtagaataacattactcat 1000
Qy 814 gttcaaaagacctctgtgtgctgctctaataatgtagtgactgttttccctaaaggagtgt 873
Db 1001 gttcaaaagacctctgtgtgctgctctaataatgtagtgactgttttccctaaaggagtgt 1060
Qy 874 tctggcccaggggattctgtgaacaggctgggaagcattctcaagattcttccagggttata 933
Db 1061 tctggcccaggggattctgtgaacaggctgggaagcattctcaagattcttccagggttata 1120
Qy 934 cttactagcacacagatgattcattacagagtgagtgatttatttcaatcaacatctctcagt 993
Db 1121 cttactagcacacagatgattcattacagagtgagtgatttatttcaatcaacatctctcagt 1180
Qy 994 gtccttgcccatactgaaattcatttcccaatttcccaatttggcccaattctcaagacctcaaat 1053
Db 1181 gtccttgcccatactgaaattcatttcccaatttggcccaatttctcaagacctcaaat 1240
Qy 1054 gtcattccatttaataatcacaggattaaacttttttttaacctgggaagaattcaatgtta 1113
Db 1241 gtcattccatttaataatcacaggattaaacttttttttaacctgggaagaattcaatgtta 1300
Qy 1114 catgagctatgggaatttaattacataattttgttttccagtgcaaaagatgactaagctcc 1173
Db 1301 catgagctatgggaatttaattacataattttgttttccagtgcaaaagatgactaagctcc 1360
Qy 1174 tttaacctcccccttgttggatttttttccagataaaagttaaagtcttagccttgt 1233
Db 1361 tttaacctcccccttgttggatttttttccagataaaagttaaagtcttagccttgt 1420
Qy 1234 actgaggtgtatacagacagcctctcccaatccctccagccttatctgtcatcaccat 1293
Db 1421 actgaggtgtatacagacagcctctcccaatccctccagccttatctgtcatcaccat 1480
Qy 1294 caacccctcccatnysacctaaacaaataacttgaatttcttgaactgtcaggnc 1353
Db 1481 caacccctcccatcaccacaaacaaataacttgaatttcttgaactgtcaggac 1540
Qy 1354 atacattttctctgtcgtgagaagctcttctgtcttcttaantctagaatgatgtaa 1413
Db 1541 atacattttctctgtcgtgagaagctcttctgtcttcttaantctagaatgatgtaa 1600
Qy 1414 agttttgaataagttgactatcttactctatgcaagaaggacacatatgagattcatc 1473

Db 1601 agttttgaataagttgactatcttacttctatgcaagaaggacacatatgagattcatc 1660
Qy 1474 atcacatgagacagacaaataactaaagtgaatttgattataaagatttagataaata 1533
Db 1661 atcacatgagacagacaaataactaaagtgaatttgattataaagatttagataaata 1720
Qy 1534 tgaatatgaaagacacacagagaggaatttttatggggcacgctttgtaagcctgggatgga 1593
Db 1721 tgaatatgaaagacacacagagaggaatttttatggggcacgctttgtaagcctgggatgga 1780
Qy 1594 agmaagcagcagggaacctcatagttctttatataataataacttctctctctatc 1653
Db 1781 agcaagcagcagggaacctcatagttctttatataataataacttctctctctatc 1840
Qy 1654 acaatatccaaacagcttttcaacaagaattctgagtgcaaatcccaaaagtaaccttt 1713
Db 1841 acaatatccaaacagcttttcaacaagaattctgagtgcaaatcccaaaagtaaccttt 1900
Qy 1714 atccatttcatgtgtgagtcgcttttagaatttttggcaaatcacttggcacttatctca 1773
Db 1901 atccatttcatgtgtgagtcgcttttagaatttttggcaaatcacttggcacttatctca 1960
Qy 1774 acttggagatgtgttctgtctgttagttaaattgaaagaataaggcactctctgtagcaca 1833
Db 1961 acttggagatgtgttctgtctgttagttaaattgaaagaataaggcactctctgtagcaca 2020
Qy 1834 ctttaggttctactcctgtgcaataaagaatttacaaga 1872
Db 2021 ctttaggttctactcctgtgcaataaagaatttacaaga 2059
RESULT 4
AAAA06687
ID AAA06687 standard; cDNA; 3112 BP.
XX
AC AAA06687;
XX
DT 13-JUN-2000 (first entry)
XX
DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:468.
XX
KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW immunogenic; cytostatic; vaccine; ss.
OS Homo sapiens.
XX
PN WO200004149-A2.
XX
PD 27-JAN-2000.
XX
PF 14-JUL-1999; 99WO-US15838.
XX
PR 14-JUL-1998; 98US-0115453.
PR 14-JUL-1998; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX
(CORI-) CORIXA CORP.
PI Dillon DC, Harlocker SL, Yugu J, Xu J, Mitcham JL;
XX
DR WPI; 2000-171268/15.
XX
PT New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumor protein -
XX
PS Claim 1; Page 259-260; 263pp; English.
XX
CC The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (pTP). The polypeptides

CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate antibodies
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the polynucleotides encoding the polypeptides can be used as a probe or
CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC AAY82000 to AAY82020 represent sequences used in the exemplification of
CC the present invention.

XX
SQ Sequence 3112 BP; 975 A; 587 C; 624 G; 926 T; 0 other;

```
Query Match          92.9%; Score 1739.2; DB 21; Length 3112;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1771; Conservative 4; Mismatches 4; Indels 3; Gaps 3;

QY 92  tacagaggtagaataaagaaggctgctgactttaccatctgagggccacacatctgctg 151
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1307  taataggtagaataaagaaggctgctgactttaccatctgagggccacacatctgctg 1366

QY 152  aaatggagataataacatcaactagaaacagcaagatgacataataatgctaaagtatg 211
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1367  aaatggagataataacatcaactagaaacagcaagatgacataataatgctaaagtatg 1426

QY 212  acatgttttgcacatttccagccctttaaataatccacacacaggaagcacaagaag 271
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1427  acatgttttgcacatttccagccctttaaataatccacacacaggaagcacaagaag 1486

QY 272  aagcacagagatccctgggagaaatgccggcgccatcttgggtcatcgatgagcctcg 331
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1487  aagcacagagatccctgggagaaatgccggcgccatcttgggtcatcgatgagcctcg 1546

QY 332  cccgtgctgctgcctgctgtgagggaaggacattagaaaaatgaattgatgttccctt 391
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1547  cccgtgctgctgcctgctgtgagggaaggacattagaaaaatgaattgatgttccctt 1606

QY 392  aaagatggcgagaaacagatcctctgtgtgattatttatttgacgggattacagatt 451
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1607  aaagatggcgagaaacagatcctctgtgtgattatttatttgacgggattacagatt 1666

QY 452  tgaatgaagtccacaaagtgcattaccatgcagagaaacagacagagaaatcttga 511
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1667  tgaatgaagtccacaaagtgcattaccatgcagagaaacagacagagaaatcttga 1726

QY 512  tggcttcacagacatgcacaaacaaataatggaataactgtgatgacatgaggcagccaag 571
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1727  tggcttcacagacatgcacaaacaaataatggaataactgtgatgacatgaggcagccaag 1786

QY 572  ctggggagagagataaccacggcgagaggtcaggattctgcccctgctgcctaaactgt 631
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1787  ctggggagagagataaccacggcgagaggtcaggattctgcccctgctgcctaaactgt 1846

QY 632  gcgttcataaccaaatacttcatatttcaaccctcaaaacaaagctgtgtgaatatct 691
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1847  gcgttcataaccaaatacttcatatttcaaccctcaaaacaaagctgtgtgaatatct 1906

QY 692  gatctcaggttccctctggggcccaacattctcaataatccagccacacatcttctta 751
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1907  gatctcaggttccctctggggcccaacattctcaataatccagccacacatcttctta 1966

QY 752  atatttagtccagatctgactgtgaccttctacactgtagaataacattactcatt 811
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1967  atatttagtccagatctgactgtgaccttctacactgtagaataacattactcatt 2026

QY 812  ttgttcaagaccctctgctgtgtgcctcaatatgtagctgactgttttccctaaaggagt 871
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2027  ttgttcaagaccctctgctgtgtgcctcaatatgtagctgactgttttccctaaaggagt 2086

QY 872  gttctggcccaaggatctgtgaacaggctgggaagcatctcaagatcttccaggggtta 931
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 5
AAA06689/C
ID AAA06689 standard; cDNA; 2426 BP.
XX
AC AAA06689;
XX

```
Db 2087  gttctggcccaaggatctgtgaacaggctgggaagcatctcaagatcttccagggtta 2146

QY 932  tacttactagcacacagcatgatcattacgagtgagtaattatcaataacatcatctcca 991
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2147  tacttactagcacacagcatgatcattacgagtgagtaattatcaataacatcatctcca 2206

QY 992  gtgtcttgcccatactgaaatcatttcccacttttggccatttctccaaacacctcaaa 1051
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2207  gtgtcttgcccatactgaaatcatttcccacttttggccatttctccaaacacctcaaa 2266

QY 1052  atgtcattccattaatatcacaggatgaacttttttttaacctgggaagaattcaatgt 1111
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2267  atgtcattccattaatatcacaggatgaacttttttttaacctgggaagaattcaatgt 2326

QY 1112  tacatgcagctatgggaatttaattacatattttgttttccagtgcaagatgactaaat 1171
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2327  tacatgcagctatgggaatttaattacatattttgttttccagtgcaagatgactaaat 2386

QY 1172  cctttatccctccctctgtgttgatttttttccagttaaaagttaaaatgcttagcctt 1231
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2387  cctttatccctccctctgtgttgatttttttccagttaaaagttaaaatgcttagcctt 2446

QY 1232  gtaactgagctgtatacag-cacagcctctcccatccctccagccttctctgtcatcac 1290
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2447  gtaactgagctgtatacagcagcctctcccatccctccagccttctctgtcatcac 2506

QY 1291  catcaaccctcccatnysacctaaacaaatctaaactgttaattccttgaacatgtcag 1350
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2507  catcaaccctcccatg-cacctaaacaaatctaaactgttaattccttgaacatgtcag 2565

QY 1351  gncatacatrttccctctgctgagaagctcttccctgtctcttcttaantctagaatgatg 1410
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2566  g-catacatatttccctctgctgagaagctcttccctgtctcttcttaantctagaatgatg 2624

QY 1411  taaagtttgaataagttgactatcttactcatgcaaaagaggggacacatagatc 1470
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2625  taaagtttgaataagttgactatcttactcatgcaaaagaggggacacatagatc 2684

QY 1471  atcatcacatgagacagcaaaataactaaaagtgaattgtattatagaagtttagataaat 1530
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2685  atcatcacatgagacagcaaaataactaaaagtgaattgtattatagaagtttagataaat 2744

QY 1531  atatgaatgcagaakccacagaggaatgtttatggggcagcttggtaagcctgggatg 1590
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2745  atatgaatgcagaagccacagaggaatgtttatggggcagcttggtaagcctgggatg 2804

QY 1591  tgaagmaaaggcaggaacacctcatagatcttctataataatacttcttctctctct 1650
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2805  tgaagcaaaggcaggaacacctcatagatcttctataataatacttcttctctctct 2864

QY 1651  atcaaatatccaaacaaagcttttccagaaatcagcagtgcaaatccccaaaggtaaac 1710
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2865  atcaaatatccaaacaaagcttttccagaaatcagcagtgcaaatccccaaaggtaaac 2924

QY 1711  ttatccatttcatggtgagtgcgcttttagaattttggcaaatcatcactggtcacttacc 1770
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2925  ttatccatttcatggtgagtgcgcttttagaattttggcaaatcatcactggtcacttacc 2984

QY 1771  tcaactttgagatgtgtttgtccttgtagtttaattgaaagaataagggcactcttgtgag 1830
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2985  tcaactttgagatgtgtttgtccttgtagtttaattgaaagaataagggcactcttgtgag 3044

QY 1831  ccactttagggttccactcctcgcaataaaagaatttacaaga 1872
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3045  ccactttagggttccactcctcgcaataaaagaatttacaaga 3086
```

[illegible]

	Query Match	92.6%	Score 1734;	DB 21;	Length 2426;
	Best Local Similarity	98.4%;	Pred. No. 0;		
	Matches 1776;	Conservative 4;	Mismatches 21;	Indels 3;	Gaps 3;
Qy	65	atctgcgatgtggaagacctgatgacagaggtgagaataagaagcgtcgtgact	124		
Db	1802	ATTTGTCCTCAATGGTCGTATGATTTCCAGGTGAGAAATAGAAGGCTGCTGACT	1743		
Qy	125	ttaccatctgaggccacacatctgctgaatggagataattaaacatcactagaacagca	184		
Db	1742	TTACCATCTGAGGCCACACATCTGCTGAATGGAGATAATTAAACATCACTAGAAACAGCA	1683		
Qy	185	agatgacaataataatgtctaagtagtgacatgtttttgcacatttccagccctttaaat	244		
Db	1682	AGATGCAATAATAATGCTAAAGTAGTGACATGTTTTTCACATTTTCCAGCCCCTTTAAAT	1623		
Qy	245	atccacacacaggaagcacaaaaggaagcacagagatccctgggagaaatgcccgccc	304		
Db	1622	ATCCACACACACGAGACGACAAAAGGAGGACACAGAGATCCCTGGGAGAAATGCCCGGCC	1563		
Qy	305	gccatcttgggtcatcgatgagcctcgccctgtgccttggtcccgcttctgagggaagac	364		

QY 1444 tgcaagaaggagacacatatgagattcatcatcacatgagacagcaaaactactaaagtgt 1503
Db 424 TGCAGAGAGGACACATATGAGATTTCATCATCATGACACAGCAAACTAAAGAGTGT 365
QY 1504 aatttgattataagagtttagataataatatataatgcgaagkccacagaggaatgttt 1563
Db 364 AATTGATTATAGAGTTTAGATATAATATGAAATGCAAGAGCCACAGAGGAATGTTT 305
QY 1564 atggggcagctgttaagctggatgtgaagaaaggcaggaacccatcatagtattcta 1623
Db 304 ATGGGGCAGCTTGTGAAGCTGGGATGTGAAGCAAGGAGGAACTCATAGTATCTTA 245
QY 1624 tataataacttcatctctatctctatcacataatccacaagaagcttttccacagaattc 1683
Db 244 TATAATATACTTCATTCTCTATCTATCAACAATCAACAAGCTTTTCACAGAATTC 185
QY 1684 atgcagtcgaatcccccaagtaaacctttatccattcatcattgagtcgctttagaat 1743
Db 184 ATGCAGTCGAATCCCAAGGTAACCTTTATCCATTTTCATGAGTGGCTTTAGAAAT 125
QY 1744 ttggcaaatcatctgctcattctctcaactttgagatgtgtttgctctgtagttaa 1803
Db 124 TTGGCAATCATCTGGTGGTCTCTCACTTTTGAGATGTGTTGTCTCTGTAGTTAA 65
QY 1804 ttgaagaataaggcactctgtgagccacttttaggttcaactcctcctggcaataaagaat 1863
Db 64 TTGAAGAATAAGGGCAGCTTGTGAGCCACTTTAGGGTTTCACTCTCGCAATAAAGANT 5
QY 1864 ttac 1867
Db 4 TTAC 1.

RESULT 6
AA06688/c
ID AA06688 standard; cDNA; 2229 BP.
XX
AC AA06688;
DT
XT 13-JUN-2000 (first entry)
XX
DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:469.
XX
KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW immunogenic; cytostatic; vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200004149-A2.
XX
PD 27-JAN-2000.
XX
PF 14-JUL-1999; 99WO-US15838.
XX
PR 14-JUL-1998; 98US-0115453.
PR 14-JUL-1998; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX
XX (CORI-) CORIXA CORP.
XX
XX Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;
PI WPI; 2000-171268/15.
XX
XX New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumor protein -
XX
PS Claim 1; Page 260-261; 263pp; English.

XX The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate antibodies
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the polynucleotides encoding the polypeptides can be used as a probe or
CC to modulate the expression of the polypeptides. AA06241 to AA06691 and
CC AA082000 to AA082020 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 2229 BP; 654 A; 447 C; 481 G; 647 T; 0 other;

Query Match 91.6%; Score 1715; DB 21; Length 2229;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1779; Conservative 4; Mismatches 21; Indels 5; Gaps 5;

QY 65 atctgcattggtgggaagcctgtagatcacagaggtgagaataagaaggctgtgact 124
Db 1808 ATTGTCTCTCTCAAAATGGTCTGATGATTTCCAGGTGAGAAATAGAAAGGCTGCTGACT 1749
QY 125 ttaccatctgagccacacacatctgctgaatgagataattaacatcacataaagacga 184
Db 1748 TTACCATCTGAGGCCACACATCTGCTGAATGGAGATAATTACATCATTAGAACAGCA 1689
QY 185 agatgacaataatgtctaaagttagtgacatgtttttgcacattttccagcccttttaaa 244
Db 1688 AGATGACAATAATAGTCTTAAGTAGTAGACATGTTTTGACATTTCCAGCCCTTTAAAT 1629
QY 245 atccacacacaggaagcaaaaaggacagacagatccctgggagaaaatgcgcgccc 304
Db 1628 ATCCACACACAGGAAGCACAAAAGGAGACACAGAGATCCCTGGGAGAAATGCCCGGCC 1569
QY 305 gccatctgggtcatcgtgagcctcgccctgctgctgctgctgctgctgctgctgctg 364
Db 1568 GCCATCTTGGGTCATCGATGACGCTCGCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 1509
QY 365 attagaaaatgaattgattgtgttccttaaggatgggcaggaaaacagatcctgtgttg 424
Db 1508 ATTAGAAAATGAATTGATGTGTCTTAAAGATGGGAGGAGGAGAAACAGATCCTGTGTGG 1449
QY 425 atatttattgaacgggattacagattgaaatgaatgaatgaatgaatgaatgaatga 484
Db 1448 ATATTATTGTAACGGGATTTACAGATTTGAAATGAAGTCACAAAGTGAGCATTTACCAATG 1389
QY 485 agaggaacacagacagaaaaatcttgatggcttcacaagacatgcaacaaacaaatgga 544
Db 1388 AGAGGAACACACAGAGAAAATCTTGATGGCTTCACAGACATGCAACAAACAAATGGA 1329
QY 545 atactgtgatgacatgagcagcgaagcctgggagagagataaccacggcgagaggtgca 604
Db 1328 ATACTGTGATGATGAGGACGCCAAGCTGGGGAGGAGATAACCAGGGGACAGGGGTCA 1269
QY 605 ggattctggcctgctgctaaactgctgctgctgctgctgctgctgctgctgctgctg 564
Db 1268 GGATTCTGGCCCTGCTGCTAAACTGCGGTTTCATAAACCAAAATCATTTTCATTTCTAAC 1209
QY 665 cctcaaaaacaaagctgttgtaatactctgctcaggtctcctctctgagcccaaatct 724
Db 1208 CCTCAAAACAAAGCTGTGTGTAATATCTGATCTCTAGGTTTCTCTGCGCCCAACATCT 1149
QY 725 ccatatccagccacactcatttttaattatttagtccagatctgtactgtgaccttt 784
Db 1148 CCATATATCCAGCCACACTCAATTTTAAATATTTAGTTCCACAGATCTGTACTGTGACCTTT 1089
QY 785 ctacactgtagaataaacattactacttttttcaagaccctctgtgtgtgtgtgtgtgt 844
Db 1088 CTACACTGTAGAATAACATTTACTCTATTTTGTTCAAAGAGCCCTTCGTTGCTGCTTAATA 1029

QY 845 tgaagtgactgttttcttaaggagtggtctgtgcccaggggatctgtgaacaggtggg 904
Db 1028 TGTAAGTACTGCTTTTCTTAAGGAGTGTCTGCGCCAGGGGATCTGTGAACAGGCTGGG 969
QY 905 aagcatcgaagatcttccaggggttactactactagcacagcatgatcatcagag 964
Db 968 AAGCATCTCAAGATCTTCCAGGGTTATACTTACTAGCACACAGCATGATCATACGGAG 909
QY 965 tgaattactaataacatcatcctcagtgctcttggcccatactgaaattcattcccaac 1024
Db 908 TGAATATCTAATCAACATCATCTCCAGTGTCTTGGCCCATACTGAAATTCATTTCCAC 849
QY 1025 ttttggccattctcaagacacctcaaatgtctattccattcaatacacaggaattacatt 1084
Db 848 TTTTGTGCCATCTCAAGACCTCAAAATGTCTATCCATTAATATACAGGATTAACT-TT 790
QY 1085 ttttttaacctggaagaattcaatgtttacatgcagctatggaatttaattacattatt 1144
Db 789 TTTTAACTTGAAGTGAAGAAATTCATGTTTACATGCAGCTATGGGAATTTAATTACATATT 730
QY 1145 tttttccagtgcaagatgactaagctctttatccctccctcttctgtttgttttttttc 1204
Db 729 TGTTTTCAGTCAAGAGATGATAAGTCTTTATCCCTCCCTTTGTGTGATTTTTC 670
QY 1205 cagtataaagttaaaatgcttagcctgttactgaggtgtgtatcacg-cacagcctctccc 1263
Db 669 CAGTATAAAGTTAAATGCTTAGCCTTGTACTGAGGCTGTATACAGCCACACGCTCTCCC 610
QY 1264 catcctccagccttatctgtcatcacaccatcaaccctccccaatnysacctaaacaaatc 1323
Db 609 CATCCCTCCAGCTTATCTGTATCACCATCAACCCCTCCCATG-CACCTAAACAAATC 551
QY 1324 taactgttaattccttgaaatgtcagncatacaattttccttctgctgagaagctct 1383
Db 550 TAACTGTGAATTTCTTGAACTGTCAGG-CATACATATTATCTCTGCTGAGAAGCTCT 492
QY 1384 tcttctctcttaactcagaatgagtaaaagtcttgaatgaagtgactatctacttca 1443
Db 491 TCCTGTCTCTTAATCTAGATGATGATAAGTTTTGTATAAGTTGACTATCTTACTTCA 432
QY 1444 tgcaagaaggagacacatgatgatctcatcacatgagacagcaaatactaaaagtgt 1503
Db 431 TGCAAGAAGGGACACATATGAGATTCATCATCATGAGACAGCAAAATACTAAAAAGTGT 372
QY 1504 aatttgattataagatttagataaataatgaaatgcaagkccacagaggggaattgttt 1563
Db 371 -ATTTGATTATAAGAGTTTAGATAAATATATGAATGCAAGGCCACAGAGGGAATGTTT 313
QY 1564 atggggcacgtttgtaagcctggagtgtaagmaaaggcagggaacctcatagtatctta 1623
Db 312 ATGGGGCACGTTTGTAGCCTGGGATGTGAAGCAAAAGGCAGGGAACCTCATATATCTTA 253
QY 1624 tataataacttcatcttctctatctctatcacaaatccacaagaagcttttcagaatc 1683
Db 252 TATAATATACTTCAATTTCTCTATCTATCAATATCCAAAGCTTTTTCACAGAATTC 193
QY 1684 atgcagtgcaaatcccaagtaacctttatccatttcagtgtagtcgctttagaat 1743
Db 192 ATGCAGTGCAAATCCCAAGGTAACCTTTATCCATTTTCATGGTGAGTGCCTTTAGAAAT 133
QY 1744 ttggcaaatcatctactgggtcacttatctcaactttgagatgtgtttgtcctgtagttaa 1803
Db 132 TTTTGCMAATCATATGCTGTCATCTATCTCAACTTTGAGATGTGTGCTCTGTAGTTAA 73
QY 1804 ttgaagaagaataggcactcttctgtagccacttttaggggttcaactcctggcaataagaat 1863
Db 72 TTTGAAAGAAATAGGGCACTCTTTGTGAGCCACTTTTAGGGTTCACTCCTGTGCAATAAAGAAAT 13
QY 1864 ttacaaaga 1872
Db 12 TTACAAAGA 4

RESULT 7
AAV62429
ID AAV62429 standard; cDNA; 820 BP.
XX
AC AAV62429;
XX
DT 30-DEC-1998 (first entry)
XX
DE Prostate cancer antigen (PCA3) cDNA splice variant 3.
XX
KW Prostate cancer antigen cDNA splice variant 3; PCA3; prostatic cancer;
KW PC; ds.
XX
OS Homo sapiens.
XX
PN WO9845420-A1.
XX
PD 15-OCT-1998.
XX
PF 09-APR-1998; 98WO-CA00346.
PR 10-APR-1997; 97US-0041836.
XX
PA (DIAG-) DIAGNOCURE INC.
XX
PI Bussemakers MJG;
XX
DR WPI; 1998-568347/48.
XX
PT New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
PT prevention and treatment of prostatic cancer
XX
PS Claim 4; Pages 77-78; 11pp; English.
XX
CC The present sequence represents the prostate cancer antigen (PCA3)
CC cDNA splice variant 3 sequence comprising of exons 1, 3, and 4a
CC of the PCA3 gene. The PCA3 cDNA splice variant 3 sequence,
CC isolated from a human primary prostatic tumour tissue cDNA library,
CC was found in approximately 15% of the cDNA clones isolated. The
CC invention claims for PCA3 cDNA variants and the proteins they encode.
CC The invention also claims for antibodies against PCA3 protein. The
CC antibodies are claimed to be useful for detecting PCA3 protein in
CC immunoassay tests, for diagnosing, assessing and prognosing of
CC prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin
CC or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed
CC to be useful for treating PC, while determining elevated levels of
CC PCA3 (as RNA or protein) is useful for detecting a predisposition
CC to development of PC, e.g. in prenatal tests. Detecting PCA3 protein
CC allows differentiation between malignant and benign prostatic disease,
CC and the level of PCA3 expression allows correlation with the grade of
CC tumour. PCA3 protein and its fragments are also claimed to be useful
CC in vaccines for preventing PC; in drug screens for identifying
CC specific (antagonists (potentially useful therapeutically) and for
CC studying protein-DNA interactions.
XX
SQ Sequence 820 BP; 262 A; 169 C; 191 G; 198 T; 0 other;

Query Match 43.8%; Score 820; DB 19; Length 820;
Best Local Similarity 100.0%; Pred. No. 1.9e-225;
Matches 820; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 aagaactggtcatcagaaaaaacagagggagatttgtgtgctgcagcgaggagaccag 60
Db 1 aagaactggtcatcagaaaaaacagagggagatttgtgtgctgcagcgaggagaccag 60
QY 61 gaagatctctgctggtgggaaggacctgatgatacacagagtgagaaataagaaggtgct 120
Db 61 gaagatctctgctggtgggaaggacctgatgatacacagagtgagaaataagaaggtgct 120
QY 121 gactttaccatctgagggccacacatctctgtaaatggagataatttaacatcactagaac 180
Db 121 gactttaccatctgagggccacacatctctgtaaatggagataatttaacatcactagaac 180


```
XX Homo sapiens.
OS WO200004149-A2.
PN 27-JAN-2000.
PD 14-JUL-1999; 99WO-US15838.
PF 14-JUL-1998; 98US-0115453.
PR 14-JUL-1998; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX (CORI-) CORIXA CORP.
PA Dillion DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;
XX WPI; 2000-171268/15.
XX New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumor protein .
XX Claim 1; Page 199-200; 263pp; English.
XX The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate antibodies
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the polynucleotides encoding the polypeptides can be used as a probe or
CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC AAY82000 to AAY82020 represent sequences used in the exemplification of
CC the present invention.
XX Sequence 718 BP; 222 A; 145 C; 169 G; 172 T; 10 other;
SQ
Query Match 27.4%; Score 513.2; DB 21; Length 718;
Best Local Similarity 97.3%; Pred No. 2.3e-137;
Matches 585; Conservative 0; Mismatches 9; Indels 7; Gaps 6;
QY 27 ggagattgtgtgg-ctgcagccgagggagaccaggaagatctgcatgtgtggaagacc 85
Db 1 ggagattgtgtgttcagccgagggagaccaggaagatctgcatgtgtggaagacc 60
QY 86 tgatgatacagaggtgagaaataagaagctgctgaactttaccatctgagccacacat 145
Db 61 tgatgatacagaggtgagaaataagaagctgctgaactttaccatctgagccacacat 120
QY 146 ctgctgaaatgagataatatacatcactagaaacagcaagatgacataataatgtctaa 205
Db 121 ctgctgaaatgagataatatacatcactagaaacagcaagatgacataataatgtctaa 180
QY 206 gtatgacatgttttgcacatttccagcccttttaaatatccacacacaggaagcac 265
Db 181 gtatgacatgttttgcacatttccagcccttttaaatatccacacacaggaagcac 240
QY 266 aaaagaacacagagatccctgggagaataatcccgccgcatcttgggtcatgatga 325
Db 241 aaaagaacacagagatccctgggagaataatcccgccgcatcttgggtcatgatga 300
QY 326 gctcgcctgtgcctggtccgctgtgaggaagacattagaaaaatgaattgatgtg 385
Db 301 gctcgcctgtgcctgntcccgctgtgaggaagacattagaaaaatgaattgatgtg 360
QY 386 ttccctaaagatggcaggaagaaacagatccctgtgtgtgatatattttgaacgggatta 445
Db 361 ttccctaaagat-ggcaggaagaaacagatccctgtgtgtgatatattttgaacgggatta 419
QY 446 cagatttgaatgaagtcaaaagtgcacaaattaccattaccatgagaggaagaaacagacgagaaa 505
Db 420 cagatttgaatgaagtcaaaagtgcacaaattaccattaccatgagaggaagaaacagacgagaaa 479
QY 506 tcttgatgcttcacaaagacatgcaacaaacaaataatggaatactgtgatgacatgagcca 565
Db 480 tcttgatgg-ttcaacaagacatgcaacaaacaaataatggaatactgtgatgacacag--c 536
QY 566 gccaaactgggagagagataaacacagggcagagggctcaggattctggccctgctccta 625
Db 537 agccaactgggagagagat-accacggggcaga-ggtcaggattctggccctgctccta 594
QY 626 a 626
Db 595 a 595
RESULT 11
AAC06768
ID AAC06768 standard; cDNA; 437 BP.
XX AAC06768;
XX 06-OCT-2000 (first entry)
XX Human secreted protein 5' EST, SEQ ID NO: 10843.
DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping; ss.
XX Homo sapiens.
OS EP1033401-A2.
PN 06-SEP-2000.
PD 21-FEB-2000; 2000EP-0200610.
PF 26-FEB-1999; 990S-0122487.
PR (GEST ) GENSET.
PA Dumas Milne Edwards J, Duclert A, Giordano J;
PI WPI; 2000-500381/45.
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX Claim 1; SEQ ID 10843; 71pp + CD-ROM; English.
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors.
XX Sequence 437 BP; 140 A; 87 C; 118 G; 88 T; 4 other;
SQ
```



```

Query Match      22.1%; Score 414.2; DB 21; Length 437;
Best Local Similarity 99.0%; Pred. No. 4.8e-109;
Matches 413; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 agaagctgcatcagaaaaacagagggagattgtgtgctgcagccgagagaccag 60
DB 21 agaagctgcatcagaaaaacagagggagattgtgtgctgcagccgagagaccag 80
QY 61 gaagatctcatgttggaagacacctgatgatacagagtgagaataaagaagctgct 120
DB 81 gaagatctcatgttggaagacacctgatgatacagagtgagaataaagaagctgct 140
QY 121 gactttaccatctgagccacacatctgctgaaatggagataattaacatcactagaac 180
DB 141 gactttaccatctgagccacacatctgctgaaatggagataattaacatcactagaac 200
QY 181 agcaagatgacaataatgtctaagtagtgacatgtttttgcacatttccagcccttt 240
DB 201 agcaagatgacaataatgtctaagtagtgacatgtttttgcacatttccagcccttt 260
QY 241 aaatatccacacacagagaagacacaaaaggaagcacagatccctggagaaatgcc 300
DB 261 aaatatccacacacagagawnacaaaaggaagcacagatccctggaganaatgcc 320
QY 301 ggcgcgcattctgggtcatcgatgagctcgccctgtgctgtgctccgcttgtgaggaa 360
DB 321 ggcgcgcattctgggtcatcgatgagctcgccctgtgctgtgctccgcttgtgaggaa 380
QY 361 gacattagaaaatgaattgatgtgttccttaagatgggcaggaacacagatcct 417
DB 381 ggcattagaaaatgratgtgtgttccttaagatgggcaggaacacagatcct 437

RESULT 12
AAZ33445
ID AAZ33445 standard; cDNA; 359 BP.
XX
AC AAZ33445;
XX
DT 08-DEC-1999 (first entry)
XX
DE Human prostate cancer-associated EST 23.
XX
KW Expressed sequence tag; EST; prostate tumor; antitumor; treatment;
KW gene therapy; tissue specificity human; ss.
XX
OS Homo sapiens.
XX
PN DE19811193-A1.
XX
PD 16-SEP-1999.
XX
PF 10-MAR-1998; 98DE-1011193.
XX
PR 10-MAR-1998; 98DE-1011193.
XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;
DR WPI; 1999-519628/44.
DR P-PSDB; AAY48243.
XX
PT New nucleic acid expressed at high level in prostatic tumor tissue and
PT encoded polypeptides, useful for treating cancer and screening for
PT therapeutic agents
XX
PS Claim 1a; 87; 166pp; German.
XX
CC This invention describes novel nucleic acid sequences (A) that are
CC expressed at high level in prostatic tumor tissue and encode gene
CC products or their fragments. The products of the invention have

```

```

CC antitumor activity. Polypeptides (I) encoded by (A) are used: (i) for
CC identifying agents for treatment of prostatic cancer and (ii) for
CC therapy of prostate cancer, optionally where expressed by gene therapy
CC methods. (A) is also used to isolate full-length genes (for gene therapy)
CC and for recombinant production of (I), which can be used to raise
CC specific antibodies. (A) are identified by assembly of ESTs (expressed
CC sequence tags) before they are analyzed for expression pattern (tissue
CC specificity). This approach eliminates many of the false results, as
CC regards tissue specificity, associated with known methods that use
CC single (usually short) ESTs. AAZ33423-233476 represent expressed
CC sequence tags described in the method of the invention.
XX
SQ Sequence 359 BP; 121 A; 75 C; 94 G; 69 T; 0 other;

```

```

Query Match      17.6%; Score 330; DB 20; Length 359;
Best Local Similarity 100.0%; Pred. No. 6.5e-85;
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaagctgcatcagaaaaacagagggagattgtgtgctgcagccgagagaccag 60
DB 30 agaagctgcatcagaaaaacagagggagattgtgtgctgcagccgagagaccag 89
QY 61 gaagatctcatgttggaagacacctgatgatacagagtgagaataaagaagctgct 120
DB 90 gaagatctcatgttggaagacacctgatgatacagagtgagaataaagaagctgct 149
QY 121 gactttaccatctgagccacacatctgctgaaatggagataattaacatcactagaac 180
DB 150 gactttaccatctgagccacacatctgctgaaatggagataattaacatcactagaac 209
QY 181 agcaagatgacaataatgtctaagtagtgacatgtttttgcacatttccagcccttt 240
DB 210 agcaagatgacaataatgtctaagtagtgacatgtttttgcacatttccagcccttt 269
QY 241 aaatatccacacacagagaagacacaaaaggaagcacagatccctgggagaaatgcc 300
DB 270 aaatatccacacacagagaagacacaaaaggaagcacagatccctgggagaaatgcc 329
QY 301 ggcgcgcattctgggtcatcgatgagccctc 330
DB 330 ggcgcgcattctgggtcatcgatgagccctc 359

```

```

RESULT 13
AAA06520/c
ID AAA06520 standard; cDNA; 301 BP.
XX
AC AAA06520;
XX
DT 13-JUN-2000 (first entry)
XX
DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:287.
XX
KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW immunogenic; cytostatic; vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200004149-A2.
XX
PD 27-JAN-2000.
XX
PF 14-JUL-1999; 99WO-US15838.
XX
PR 14-JUL-1998; 98US-0115453.
PR 14-JUL-1998; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX

```


Qy 1567 gggcagctttagcctgggatgtagmaagcagggaacccctcatagtatcttat 1626
Db 177 WWWWWW... 118
Qy 1627 aatatacttctctctctctctctcaaatatccaaagctttttcacagaattcatg 1686
Db 117 WWWWWW... 58
Qy 1687 cagtgcaatcccaaggtaacctttatccatttcatggtgagcgcttttagaat 1743
Db 57 WWWWWW... 1

RESULT 15

AAF58254/C
ID AAF58254 standard; DNA; 936 BP.

AC AAF58254;

DT 24-APR-2001 (first entry)

DE Oligonucleotide D1875.

XX Electron-transfer group; ETM; mismatch; genotyping;

KW gene expression; ss.

XX Synthetic.

XX WO200107665-A2.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000WO-US20476.

PR 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.

XX (CLIN-) CLINICAL MICRO SENSORS INC.

XX UmeK RM;

XX WPI; 2001-159728/16.

XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface

XX Example 6; Page 127; 159pp; English.

XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.

XX Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match 7.3%; Score 136; DB 22; Length 936;

Best Local Similarity 0.9%; Pred. No. 5.9e-29;

Matches 7; Conservative 491; Mismatches 279; Indels 0; Gaps 0;

Qy 967 aattatctaatcaacatcatcctcagtgctgtcttggccctactgaataatccactt 1026
Db 777 WWWWWW... 718

Qy 1027 tigtgccattctcaagacctcaaatgtcattccatttaataatcacaggattattt 1086

Db 717 WWWWWW... 658

Qy 1087 tttttaacctggaagaattcaatggttacatgcagctatgggaatttaattacatat 1146

Db 657 WWWWWW... 598
Qy 1147 ttttccagtgcaaaagatgaactaagtcctttatccctcttctgttattttttcca 1206
Db 597 WWWWWW... 538
Qy 1207 gtaaaagttaaaatgcttagcttactgaggtgtatcacagcacagcctctcccat 1266
Db 537 WWWWWW... 478
Qy 1267 cctccagccttatctgtcctatccatcaacccctcccatnysacctaaacaaatctaa 1326
Db 477 WWWWWW... 418
Qy 1327 cttgtaattccttgacatgctcagncatcacatttctctctgcctgagaagctctcc 1386
Db 417 WWWWWW... 358
Qy 1387 ttgtctcttaantctagaatgatgtaaagttttgataagttgacttatcttactcatgc 1446
Db 357 WWWWWW... 298
Qy 1447 aaagaaggacacatatgagattcatcatcacatgagacagcaaatatactaaagtta 1506
Db 297 WWWWWW... 238
Qy 1507 ttgattataagagtttagataaataatgaaatgcaagakccacagaggggaatgttatg 1566
Db 237 WWWWWW... 178
Qy 1567 gggcagctttagcctgggatgtagmaagcagggaacccctcatagtatcttat 1626
Db 177 WWWWWW... 118
Qy 1627 aatatacttctctctctctctcaaatatccaaagctttttcacagaattcatg 1686
Db 117 WWWWWW... 58
Qy 1687 cagtgcaatcccaaggtaacctttatccatttcatggtgagcgcttttagaat 1743
Db 57 WWWWWW... 1

Search completed: July 31, 2001, 07:26.09

Job time: 17781 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 07:41:09 ; Search time 10902.7 Seconds
(without alignments)
2655.824 Million cell updates/sec

Title: US-09-402-713A-3
Perfect score: 1872
Sequence: 1 agaagctggcatcagaaaa.....caataaagaatttacaaga 1872

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_bal: *
2: gb_ba2: *
3: gb_ba3: *
4: gb_in1: *
5: gb_in2: *
6: gb_in3: *
7: gb_om: *
8: gb_ov: *
9: gb_pat1: *
10: gb_pat2: *
11: gb_ph: *
12: gb_pl1: *
13: gb_pl2: *
14: gb_pl3: *
15: gb_pl4: *
16: gb_bal: *
17: gb_ba2: *
18: gb_fun: *
19: em_htgo_hum: *
20: em_htgo_inv: *
21: em_htgo_rod: *
22: em_htg_hum1: *
23: em_htg_hum2: *
24: em_htg_hum3: *
25: em_htg_hum4: *
26: em_htg_hum5: *
27: em_htg_hum6: *
28: em_htg_hum7: *
29: em_htg_hum8: *
30: em_htg_inv1: *
31: em_htg_inv2: *
32: em_htg_other: *
33: em_htg_rod: *
34: em_hum1: *
35: em_hum2: *
36: em_hum3: *
37: em_hum4: *
38: em_hum5: *
39: em_hum6: *
40: em_hum7: *
41: em_in: *
42: em_om: *
43: em_or: *

44: em_ov: *
45: em_pat: *
46: em_ph: *
47: em_pl: *
48: em_ro: *
49: em_sts: *
50: em_sy: *
51: em_un: *
52: em_vl: *
53: gb_sts1: *
54: gb_sts2: *
55: gb_sts3: *
56: gb_sy: *
57: gb_un: *
58: gb_vl1: *
59: gb_vl2: *
60: gb_vtg1: *
61: gb_vtg2: *
62: gb_vtg3: *
63: gb_vtg4: *
64: gb_vtg5: *
65: gb_vtg6: *
66: gb_vtg7: *
67: gb_vtg8: *
68: gb_vtg9: *
69: gb_vtg10: *
70: gb_vtg11: *
71: gb_vtg12: *
72: gb_vtg13: *
73: gb_vtg14: *
74: gb_vtg15: *
75: gb_vtg16: *
76: gb_vtg17: *
77: gb_vtg18: *
78: gb_vtg19: *
79: gb_vtg20: *
80: gb_vtg21: *
81: gb_vtg22: *
82: gb_vtg23: *
83: gb_vtg24: *
84: gb_vtg25: *
85: gb_pr1: *
86: gb_pr2: *
87: gb_pr3: *
88: gb_pr4: *
89: gb_pr5: *
90: gb_pr6: *
91: gb_pr7: *
92: gb_pr8: *
93: gb_pr9: *
94: gb_ro1: *
95: gb_ro2: *
96: gb_in4: *
97: gb_pr10: *
98: em_ba3: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1772.4	94.7	3923	88	AF103907	AF103907 Homo sapi
2	1589.4	84.9	5435	88	AF103908	AF103908 Homo sapi
3	1555.8	83.1	164371	80	AL390239	AL390239 Homo sapi
4	1361.4	72.7	173831	80	AL359314	AL359314 Homo sapi
c 5	1317.8	70.4	267581	80	AL358573	AL358573 Homo sapi
6	330	17.6	359	9	AX018075	AX018075 Sequence
7	100.2	5.4	143675	79	AL161625	AL161625 Homo sapi
8	100.2	5.4	267581	80	AL358573	AL358573 Homo sapi

```
c 9 98.6 5.3 172298 61 AC009556 Homo sapi
10 98 5.2 580 89 AF279290 Homo sapi
11 59.8 3.2 7218 10 I66494 Sequence 14
12 57.4 3.1 7218 10 I66494 Sequence 14
13 52.6 2.8 1141 10 AC083744 Sequence
14 47.2 2.5 137064 74 AC073363 Homo sapi
15 47.2 2.5 152571 66 AC021589 Homo sapi
16 47.2 2.5 160133 69 AC025140 Homo sapi
17 47.2 2.5 167574 62 AC011274 Homo sapi
18 47.2 2.5 177396 74 AC069219 Homo sapi
19 47.2 2.5 189915 70 AC026861 Homo sapi
20 47.2 2.5 191034 80 AL356372 Homo sapi
21 47 2.5 10968 9 AR029638 Sequence
22 47 2.5 10968 14 LEU65391 Lycopersico
23 47 2.5 59202 13 AF220603 Lycopersi
24 47 2.5 60831 13 AF220602 Lycopersi
25 45.8 2.4 188781 67 AC022546 Homo sapi
26 45.8 2.4 323479 77 AC087779 Pan trogl
27 45.6 2.4 40873 83 CEH04109 Caenorhabd
28 45.6 2.4 42141 6 CEY54G9A Caenorhabd
29 45.6 2.4 300197 83 CEY54G9 Caenorhabd
30 45.4 2.4 5134 9 AR085492 Sequence
31 45.4 2.4 5475 9 AR029637 Sequence
32 45.2 2.4 1141 10 AC083744 Sequence
33 45 2.4 191319 63 AC015653 Homo sapi
34 44.4 2.4 151553 90 AL139342 Human DNA
35 44.4 2.4 181545 69 AC025529 Mus muscu
36 44.2 2.4 163958 4 AC010705 Drosophil
37 44 2.4 106935 6 CEY37B3 Caenorhab
38 43.8 2.3 171471 64 AC015844 Homo sapi
39 43.2 2.3 120984 86 AC006463 Homo sapi
40 43.2 2.3 184316 68 AC024022 Homo sapi
41 42.8 2.3 3269 2 AF270032 Staphyloc
42 42.8 2.3 9131 85 AB038781 Homo sapi
43 42.6 2.3 930 53 CNS06P22 T7 end of
44 42.6 2.3 8912 85 AB038782 Homo sapi
45 42.6 2.3 10708 89 AF113616 Homo sapi
```

ALIGNMENTS

```
RESULT 1
AF103907 3923 bp mRNA PRI 14-AUG-2000
LOCUS Homo sapiens non-coding RNA DD3 sequence.
AF103907
VERSION AF103907.1 GI:6165973
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3923)
Bussemakers,M.J., van Bokhoven,A., Verhaegh,G.W., Smit,F.P.,
Karthaas,H.F., Schalken,J.A., Debruyne,F.M., Ru.N. and Isaacs,W.B.
DD3: a new prostate-specific gene, highly overexpressed in prostate
cancer
Cancer Res. 59 (23), 5975-5979 (1999)
JOURNAL Cancer Res. 59 (23), 5975-5979 (1999)
MEDLINE 20072260
PUBMED 10606244
REFERENCE 2 (bases 1 to 3923)
AUTHORS Bussemakers,M.J.G., Van Bokhoven,A., Verhaegh,G.W., Smit,F.P.,
Karthaas,H.F.M., Schalken,J.A., Debruyne,F.M.J., Ru.N. and
Isaacs,W.B.
TITLE Direct Submission
JOURNAL Submitted (28-Oct-1998) Urology Research Laboratory, University
Hospital Nijmegen, P.O. Box 9101, Nijmegen 6500 HB, Netherlands
FEATURES
Location/Qualifiers
1..3923
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
```


Db 2036 TGCTGGTCCCGCTTGTGAGGGGAAGACATTTAGAAAATGAATTGATGTGTCTCTTAAAGG 2095
 QY 397 atggcagaaacacagatcctgtgtggtatattttgaacgggattacagatttgaaa 456
 Db 2096 ATGGCAGGAAACACAGATCCTGTGTGGATATTTATTTGAACGGGATTACAGATTTGAA 2155
 QY 457 tgaagtcaaaagtgcattaccatagagagagaaacagagagaaaaatccttgatggtt 516
 Db 2156 TGAAGTCAACAAGTGAAGCATTTACCAATGAGAGGAAAAACAGAGAGAAATCTTGATGGCT 2215
 QY 517 tcaagagacatgcaacaacaaatggaatactgtgatgacatgagcagcaagctggg 576
 Db 2216 TCACAGACATGCAACAACAATGGAATACCTGTGTGATGACATGAGCGACGCCAAGCTGGG 2275
 QY 577 gagagataaccacggggcagagggtcaggattctggcctgcctgcctaaactgtgcgtt 636
 Db 2276 GAGGAGATTAACACGGGCGAGAGGTCAGGATTCGGCCTGCTGCTAACTGTCGCTT 2335
 QY 637 cataaccaaatcatctcatatttctaaccctcaaaaacaaagctgttgttaatatctgatct 696
 Db 2336 CATPAACCAATCATTTTCATATTTCTAACCTCAAAAACAAAGCTGTTGTGAATATCTGATCT 2395
 QY 697 ctacgattcctctgcggcccaacattctccatatatccagccacacactcaattttaaatt 756
 Db 2396 CTACGGTTCCTCTGGGCCCAACATTTCTCCATATATCCAGCCACACTCAATTTTTTAATAT 2455
 QY 757 tagttccagatctgtactgtgacctttctacactgtagaataaacattactcattttgtt 816
 Db 2456 TAGTTCACAGATCTGTACTGTGACCTTTCTACACTGTAGAAATTAACATTACTCATTTGTT 2515
 QY 817 caaagaccttggctgtgctgctaataatgtagctgactgttttctcctaagagtggtct 876
 Db 2516 CAAAGACCTTCGTGTGTGCTGCTAATATGTAGCTGACTGTTTCTTAAAGAGTGTCT 2575
 QY 877 gcccagggatctgtgaacaggctgggaagcatctcaagatcttccagggtataactt 936
 Db 2576 GGGCCAGGGGATCTGTGAACAGCTGGGAAGCATCTCAAGATCTTTCCAGGGTTATACCT 2635
 QY 937 actagcacacagcatgatcatcagcagtgagtgaaattatctaataacatcatcctcagtgct 996
 Db 2636 ACTAGCACACAGCATGATCATTTACGGAGTGAATTAATCAACATCATCTCAGTGTCT 2695
 QY 997 ttggcccatctgaaattcatttcccaactttgtgcccattctcagacacctcaaaatgtc 1056
 Db 2696 TTTGGCCATCTAGAAATTCATTTCCCACTTTTGTGCCATTTCTCAAGACCTCAAAATGTC 2755
 QY 1057 attccattaatacacagattaaactttttttaacctgaagaattcaatgttaact 1116
 Db 2756 ATTCCATTATATACAGGATTAACCTTTTAACTTGAAGAAATTCAAATGTTACAT 2815
 QY 1117 gcagctatgggaatttaatacatattttgttttccagtgcaagatgactaagctcttt 1176
 Db 2816 GCAGCTATGGGAATTAATACATATTTTGTTCAGTGCAAGATGACTAAGTCTTT 2875
 QY 1177 atccctcccttggctgtgtatttttccagtagtaaaagttaaatgcttagcctgtact 1236
 Db 2876 ATCCCTCCCTGTTGTTGATTTTTTTCAGTATAAAGTTAAATGCTTAGCCTTGTACT 2935
 QY 1237 gaggtctgtacagcacagcctctcccatcctccagcttctctgtcatcaccatcaa 1296
 Db 2936 GAGGCTGTATACAGCACAGCCCTCTCCCATCCCTCCAGCTTATCTGTATGTCATCACCATCA 2995
 QY 1297 cccctcccatnysacctaaacaaaaactaactgttaattccttgaaactgtcaggncata 1356
 Db 2996 CCCCCTCCCATACCACCTAAACAAATCTAATCTGTAATTTCTTGAACATGTCCAGGACATA 3055
 QY 1357 catrttccctgtgctcgtgagagcttctcctgtctcttaantctagaatgatgtaaagt 1416
 Db 3056 CATATTCTCTGCTGTGAGAGCTCTCTCTGTCTCTTTAAATCTAGAATGATGATAAGT 3115
 QY 1417 ttgtaaatgtagtactatctactcatgcaaaagagagacacatatgagattcatc 1476
 Db 3116 TTTGAATAAGTTGACTATCTTACTTCATGCAAGAAAGGAGGACACATATGATGATTCATCATC 3175

QY 1477 acatgagacagcaaaataactaaaaagtgaatttgattataagagtttagataaatatga 1536
 Db 3176 ACATGAGACAGCAAAATACTAAAAGTCTAATTTTCATTATAAGAGTTTAGATAAATATATCA 3235
 QY 1537 atgcaagakccacagaggaatgtttatcgggcagctttgtaagcctgggatgtgaagm 1596
 Db 3236 AATGCAAGAGCACAGAGGGAATGTTTATGGGCGACGTTTGTAAAGCTGGGATGTGAAGC 3295
 QY 1597 aaaggcaggaacacctcatagttcttataataataataattcattcttctctctatcaca 1656
 Db 3296 AAAGGAGGGAACCTCATAGTATCTTATATAATATATATCTTCTATCTATCTATCACA 3355
 QY 1657 atatccaaacagctttttcacagaattcatgcagtgcaaaatccccaaaggttaacctttatc 1716
 Db 3356 ATATCCAAACAGCTTTTCACAGAAATTCATGCAGTGCAATCCCAAGGTAACTTTATC 3415
 QY 1717 catttcattgagtcgcttttagaattttggcaaatcactatggtcacttatctcaact 1776
 Db 3416 CATTTTCATGCTGAGTGCCTTTTAGAATTTTGGCAATCATCTGGTCACTTATCTCAACT 3475
 QY 1777 ttgaatgtgttgcctgttagttaattgaagaataaggcactctgttgagcaactt 1836
 Db 3476 TTGAGATGTTGTTGCTTGTAGTTAATTTGAAGAAATAGGCACCTCTTGTGAGCCACTT 3535
 QY 1837 tagggttactcctctggaataaagaatttacaaga 1872
 Db 3536 TAGGTTCACTCTCGCAATAAAGAATTTACAAAGA 3571

RESULT 3

AL390239

LOCUS

AL390239 164371 bp DNA HTG 09-MAR-2001
 Homo sapiens chromosome 9 clone RP11-58J3, *** SEQUENCING IN
 PROGRESS ***, 4 unordered pieces.

ACCESSION

AL390239

VERSION

AL390239.11 GI:13274794

KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Burton, J.

Direct Submission

Submitted (05-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequests@sanger.ac.uk

On Mar 12, 2001 this sequence version replaced gi:13273805.

----- Genome Center

Center: Sanger Centre

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: BA58J3

----- Summary Statistics

Sequencing program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 163337 bases at least Q40

Consensus quality: 163481 bases at least Q30

Consensus quality: 163577 bases at least Q20

Insert size: 164071; sum-of-contigs

Insert size: 166918; 4.2% error; agarose-fp

Quality coverage: 8.78x in Q20 bases; sum-of-contigs Quality

coverage: 9.07x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently

* consists of 4 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

- * This record will be updated with the finished sequence
- * as soon as it is available and the accession number will
- * be preserved.

23241	116100: contig of 100 bp
23242	116100: contig of 100 bp
23243	116100: contig of 95760 bp in length
23244	116200: gap of 100 bp
23245	116200: gap of 100 bp
23246	118243: contig of 2043 bp in length
23247	118243: gap of 100 bp
23248	118343: contig of 100 bp in length
23249	164371: contig of 46028 bp in length
23250	164371: contig of 46028 bp in length

FEATURES

source

```

1. .10437/1
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-58J3"
/clone_lib="RPC1-11.1"
1. .20240
/note="assembly_fragment:00838"
fragment_chain:1
clone_end:SP6
vector_side:left"
20341. .116100
/note="assembly_fragment:01998"
fragment_chain:1"
116201. .118243
/note="assembly_fragment:00916"
fragment_chain:1"
118344. .164371
/note="assembly_fragment:00334"
fragment_chain:1
clone_end:T7

```

BASE COUNT	50254 a	34905 c	32663 g	46249 t	300 others
ORIGIN	veccol_side: r1gn1c				

Query Match	83.1%	Score 1555.8	DB 80	Length 164371
Best Local Similarity	99.4%	Pred. No. 0		
Matches 1587	Conservative 4	Mismatches 3	Indels 3	Gaps 3
Qy 277	cagagatccctcgagagaaatgccggccgcgcacatttgggtcatcgatgagcctgcgcctg	336		
Db 33316	CACAGATCCCTGGGAGAAATCCCGCGCGCATCTTGGGTCAATGATGAGCTCGCCCTG	33375		
Qy 337	tgcctggtcccgccttctgagggaaaggacattagaaaaatgaattgatgtgttcccttaaaag	396		
Db 33376	TGCTCTGGTCCCGCTTGTGAGGGAAAGGACATTAGAAAATGAAATTGATGTGTCTCTTAAAGG	33435		
Qy 397	atggcgaggaaaaacagatccctgttgtggattatttattgaacgggattcacagatttgaaa	456		
Db 33436	ATGGCAGGAAAAACAGATCCTGTTGTGGATTATTGTGAACGGGATTACAGATTTGAAA	33495		
Qy 457	tgaagtcacaaagttagcattaccaatgagaggaaaaacagacgagaaaaatcttgatggct	516		
Db 33496	TGAAGTCACAAAGTGAGCATTTACCAATGAGAGGAAAAACAGACGAGAAAATCTTGATGGCT	33555		
Qy 517	tcacaagatgcacaacaaacaaaaatggaactctgtgatgacatgaggcagccaagctggg	576		
Db 33556	TCACAAGACATGCACAACAACAAAATGGAATACTGTGATGACATGAGGCAGGCAAGCTGGG	33615		
Qy 577	gaggagataaccacggggcagagggtcaggattctggccttgctgcctcaaaactgtgcatt	636		
Db 33616	GAGGAGATAACACGGGGCAGAGGCTCAGGATTTGGCCCTTGCTGCCTTAAACTGTGCGGTT	33675		
Qy 637	cataaccaaatcatcttcatttcttaacttctaacccctcaaaacaaaagctgtgtgaatctgatct	696		
Db 33676	CATAACCAATCATTTTCATATTCTTAACCCCTCAAAACAAGCTGTGTGAATATCTGATCT	33735		
Qy 697	ctacggttcctctcgggcccacaattctccatatatccagcccaactcatcttttaatt	756		
Db 33736	CTACGGTTCCTCTGGGGCCCAACATTCCTCATATATCCAGGCACACTCATTTTTTAATATT	33795		

Qy	757	tagttccacagatctgtactgtgacactttctacactgtagaataacataacttacttctgtt	816
Db	33796	TAGTTCCACAGATCTGTACTGTGACCTTCTACACTGTAGAATAACATTTACTCATTTTGT	33855
Qy	817	caagacccttcgtgtgtcctaaatgtagetgactgttttccctaaaggagtgctct	876
Db	33856	CAAGACCCTTTCGTGTCTGCCTTAATATGTAGCTGACTGTTTTTCCCTAAGGAGTGTTCT	33915
Qy	877	ggccagggggactctgtgaaacaggctgggaagcatctcaagatcttctccagggttatactt	936
Db	33916	GGCCAGGGGATCTGTGAACAGGCTGGCAAGCATCTCAAGATCTTTTCAGGGTTTACTTT	33975
Qy	937	actagcacacagcatgatcataagagagtgaaatatacttaatacaacatcatcctcaagtgc	996
Db	33976	ACTAGCACACAGCATGATCATTTACGGAGTGAATATCTAATCAACATCATCCTCAGTGTC	34035
Qy	997	tttgccatactgaaatcatttcccactcttctgcccactctcaagacacctcaaaatgctc	1056
Db	34036	TTTGCCCATACTGAANTTTCATTTCCACTTTTGTGCCCATCTCTCAAGACCTCAAAATGTC	34095
Qy	1057	attccattaatcacagagattaaacttttttttaacctggaagaattcaatgttacct	1116
Db	34096	ATTCCATTAAATCACAGGATTAACCTTTTTTTTTTAACCTGGAAGATTTCAATGTTACAT	34155
Qy	1117	gcagctatgggaatttaatacatatatgtttgtttccagtgcaagatgactaaagtccttt	1176
Db	34156	GCAGCTATGGGAATTTAAATACATATTTGTTTTTCCAGTGGCAAGATGACTAAGTCCCTT	34215
Qy	1177	atccctccccttgttgaattttttccagtaaaagttaaagttaaagtctagccttgact	1236
Db	34216	ATCCCTCCCCTTTGTGTTGATTTTTTTTCCAGTATAAAGTTAAATGCTTAGGCTTGTTACT	34275
Qy	1237	gaggctgtatacag-cacagcctctcccacatccctccagccttatctgtcacaaccaatca	1295
Db	34276	GAGGCTGTATACGCCACAGCCTCTCCCATCCCTCCAGCCTTATCTGTCTATCACCATCA	34335
Qy	1296	accctcccatnysacctaaacaaaatctaaactgttaactccttgacatgtcaggnact	1355
Db	34336	ACCCCTCCCATG-CACCTTAACAAAATCTAACTGTGTAATTCCTTGAACATGTGAGG-CAT	34393
Qy	1356	acattttccctctgcctgagaagctctcctgtctctctaantctagaatgatataag	1415
Db	34394	ACATTATTTCTTCTGCCCTGAGAAGCTCTTCTGTGCTCTTAANTCTAGAATGATGTAAAG	34453
Qy	1416	ttttgaataagttgactacttacttccatgcaaaagggacacatatagagattcatcat	1475
Db	34454	TTTTGAATAAGTTGACTATCTTACTTTCATGCAAAAGSGACACATATGAGATTCTATCAT	34513
Qy	1476	cacatgacagcaataactaaaaagtgtaatttgattataagagtttagataaaatatg	1535
Db	34514	CACATGAGACAGCAAAATACTAAAAGTGAATTTGATTAATAAGAGTTTAGATAAATATAG	34573
Qy	1536	aaatgcaagakccacagaggaattttatggggcacgtttgttaagcctggagatgtgaag	1595
Db	34574	AAATGCAAGAGCCACAGAGGGAATGTTTATGGGGACGTTTGTAGCCCTGGGATGTGAAG	34633
Qy	1596	maaggcgaggaaacctcatagtatcttataataataacttcattctctatctctatcac	1655
Db	34634	CAAAAGCAGGNAACCTCATGTATCTTATTAATACTTCAATTTCTTATCTCTATCTAC	34693
Qy	1656	aatatccacaagcgtttcacagaattcatgcaagtgaatcccccaagggtaacctttat	1715
Db	34694	AAATATCCAAAGCTTTTTCACAGAATTCATGCAGTGCAAAATCCCAAGGTAACCTTTAT	34753
Qy	1716	caatttcagtgtgagtgagctttagaattttggcaaatcaatactaggtcaacttatctcaac	1775
Db	34754	CAATTTTCATGTGTGAGTGCGCTTTAGAAATTTTGGCAAAATCATACTGGTCACTTATCTCAAC	34813
Qy	1776	tttgagatgtgttgccttgtagttaattgaaagaaatagggcactctcttgagccact	1835
Db	34814	TTTGAGATGTGTTTGTCTCTGTAGTTAATTTGAAGAAATAGGGCACTCTTGTAGCCACT	34873
Qy	1836	tttagggttcactctcctggcaataaagaatttacaaga	1872

```
|||||
Db 34874 TTAGGTTTCACTCTCGCAATAAGAAATTACAAAGA 34910
|||||
RESULT 4
AL359314 173831 bp DNA HTG 08-APR-2001
LOCUS Homo sapiens chromosome 9 clone RP11-108L4, *** SEQUENCING IN
DEFINITION PROGRESS ***, 3 unordered pieces.
ACCESSION AL359314
VERSION AL359314.12 GI:13396560
KEYWORDS HTG; HTGS-PHASE1; HTGS-ACTIVEFIN; HTGS-DRAFT; HTGS_FULLLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 173831)
Leongamornlert,D.
Direct Submission
Submitted (07-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Mar 20, 2001 this sequence version replaced gi:13277120.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bA108L4
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 173545 bases at least Q40
Consensus quality: 173599 bases at least Q30
Consensus quality: 173621 bases at least Q20
Insert size: 173631; sum-of-contigs
Insert size: 172123; 10.0% error; agarose-fp
Quality coverage: 10.35x in Q20 bases; sum-of-contigs Quality
coverage: 10.50x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 50595: contig of 50595 bp in length
* 50596 50695: gap of 100 bp
* 50696 157636: contig of 106941 bp in length
* 157637 157736: gap of 100 bp
* 157737 173831: contig of 16095 bp in length.
FEATURES
Source
Location/Qualifiers
1..173831
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-108L4"
/clone_lib="RPC1-11.1"
1..50595
/note="assembly fragment:04137
fragment_chain:1"
50696..157636
/note="assembly_fragment:04236
fragment_chain:1"
157737..173831
/note="assembly fragment:01656"
BASE COUNT 52107 a 36469 c 35018 g 50037 t 200 others
ORIGIN
```

```
Query Match 72.7%; Score 1361.4; DB 80; Length 173831;
Best Local Similarity 92.4%; Pred. No. 0;
Matches 1395; Conservative 4; Mismatches 107; Indels 3; Gaps 3;
QY 277 cagagatccctgggagaaatgccgcgcgcctcttgggtcatcagatgagcctgcgcctg 336
|||
Db 156237 CACAGATCCCTGGGAGAAATGCCGCGCCCATCTTGGGTCTCATCGATGAGCCTCGCCCTG 156296
QY 337 tgccttgctcccttgctgagggagagacattagaaaaatgaatgattgtgttctctaaagg 396
|||
Db 156297 TCCTCTGCTCCCTCTTGTGAGGGAAGACATTAGAAAATGAATGTCTTCTTTAAAGG 156356
QY 397 atgggcagaaacacagatcctgtgtgggataattttgaacgggattacagatttgaa 456
|||
Db 156357 ATGGCGAGAAACACAGATCCTGTGTGGGATATTTATTTCAGCGGATTTACAGATTTGAAA 156416
QY 457 tgaagtcaaaaagttagcattaccatgagagagaaacagacagaaaaatcttgatggct 516
|||
Db 156417 TGAAGTTCACAAAGTGAGCATTTACCAATGAGAGAAACAGACAGAGAAAATCTTGATGGCT 156476
QY 517 tcacaagacatcaacaacaaataaggaatactgtgatactgatactgagggagcgaagctggg 576
|||
Db 156477 TCACAAGACATGCAACAAACAAATGGAATCTGTGATGACATGAGGCAAGCAAGCTGGG 156536
QY 577 gagagataaacacaggggcagaggggtcaggattcttggccctgtgcctaaactgtgcgtt 636
|||
Db 156537 GAGGAGATAAACACCGGGGCAGAGGGTCAGGATTCGGCCCTCTGCTGCTAAACTGTGCGTT 156596
QY 637 cataaccaatcatcttcatatttctaaccctcaaaacaaagctgttgttaatactgatct 696
|||
Db 156597 CATAACCAAAATCATTTTATATTTCTTAACCTCAAAACAAAGCTGTGTGTAATATCTGATCT 156656
QY 697 ctacggttctcttggggcccaattctccatatactatccagccacacactcaatttataatt 756
|||
Db 156657 CTACGGTTCCTTCTGGGCCCAACATTTCTCCATATATCCAGCCACACTCATTTTATATAT 156716
QY 757 tagttccacagatctgtactgtgacctttctacactgtagaataaacattactcatcttctgt 816
|||
Db 156717 TAGTTCCCGAGATCTGTACTGTGACCTTTCTACACTGTAGAAATACATTTACTCATTTGTT 156776
QY 817 caaagaccttctgttctgctgctaataatgtagctgactgttttctcctaagagtgcttct 876
|||
Db 156777 CAAAGACCTTCGTGTGTGCTGCTAATATGTAGTACTGTGTGTTCCTTCCCTAAGGAGTGTCT 156836
QY 877 gccccagggatctgtgaacagcctgggaagcatctcaagatcttccaggttatactt 936
|||
Db 156837 GCCCAGGGGATCTGTGAACAGGCTGGGAAGCATCTCAAGATCTTTCCAGGCTTTATACCT 156896
QY 937 actagcacacagcatgatacttaccgagtgagtgaaattatctaatcaacatcatcctcagtgct 996
|||
Db 156897 ACTAGCACACAGCATGATCATTTACGAGTGAAATTTATCTAATCAACATCATCTCAGTGTC 156956
QY 997 ttgcccatactgaaatcatttcccaacttttggcccaacttccagcattcgaagcctcaaatgtc 1056
|||
Db 156957 TTTTGCCCACTACTGAAATTCATTTCCCACTTTTGTGCCCATCTTCAAGACCTTCAAAATGTC 157016
QY 1057 attccattaatatcacagattaaacttttttttaacctggaagaattcaatgttacct 1116
|||
Db 157017 ATTCCATTAAATATCAGGATTAACCTTTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 157076
QY 1117 gcagctatgggaatttaatacatatattttgtttccagtgcaaaagatgactgaagtccttt 1176
|||
Db 157077 GCAGCTATGGGAATTTAATTTACATATTTTGTGTTTCCAGTGCAAGATGACTAAGTCTCTTT 157136
QY 1177 atccctcccccttctgttgaatttttttccagtaataaagttaaaatgcttagcctgtact 1236
|||
Db 157137 ATCCCTCCCCCTTTGTTGATTTTTTTTCCAGTATAAAGTTAAATGCTTTAGCTTTGTTACT 157196
QY 1237 gaggtgtatcacag-cacagcctctccccatccctccagccttattctgtcatcacatca 1295
|||
Db 157197 GAGGCTGTATACAGCCACAGCCTCTCCCATCTCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 157256
```



```

* 169801 169900: gap of 100 bp
* 169901 251545: contig of 81645 bp in length
* 251546 251645: gap of 100 bp
* 251646 259029: contig of 7384 bp in length
* 259030 259129: gap of 100 bp
* 259130 262871: contig of 3742 bp in length
* 262872 262971: gap of 100 bp
* 262972 265209: contig of 2238 bp in length
* 265210 265309: gap of 100 bp
* 265310 267581: contig of 2272 bp in length.

```

FEATURES

Location/Qualifiers

```

1..267581
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-133022"
/clone.lib="RPC1-11.1"
1..25718
/note="assembly_fragment:00645
fragment_chain:1
clone_end:77
vector_side:left"
25819..29567
/note="assembly_fragment:01660
fragment_chain:1"
29668..40520
/note="assembly_fragment:03080
fragment_chain:1"
40621..50002
/note="assembly_fragment:00771
fragment_chain:1"
50103..54942
/note="assembly_fragment:02629
fragment_chain:1"
55043..59067
/note="assembly_fragment:04565
fragment_chain:1"
59168..72327
/note="assembly_fragment:00223
fragment_chain:1"
72428..79396
/note="assembly_fragment:03318
fragment_chain:1"
79497..82738
/note="assembly_fragment:00224
fragment_chain:2"
82839..87543
/note="assembly_fragment:03050
fragment_chain:2"
87644..89847
/note="assembly_fragment:01948
fragment_chain:3"
89948..92376
/note="assembly_fragment:02191
fragment_chain:3"
92477..99799
/note="assembly_fragment:00010"
99900..102117
/note="assembly_fragment:00672"
102218..109330
/note="assembly_fragment:01326"
109431..112187
/note="assembly_fragment:01498"
112288..116407
/note="assembly_fragment:01607"
116508..119386
/note="assembly_fragment:01899"
119487..121869
/note="assembly_fragment:02068"
121970..125546
/note="assembly_fragment:02469"
125647..128990
/note="assembly_fragment:02502"

```

```

misc_feature 129091..134055
/note="assembly_fragment:02669"
misc_feature 134156..138314
/note="assembly_fragment:02757"
misc_feature 138415..140612
/note="assembly_fragment:02776"
misc_feature 140713..143296
/note="assembly_fragment:03127"
misc_feature 143397..148113
/note="assembly_fragment:03634"
misc_feature 148214..151009
/note="assembly_fragment:03898"
misc_feature 151110..156616
/note="assembly_fragment:04220"
misc_feature 156717..158763
/note="assembly_fragment:04276"
misc_feature 158864..162129
/note="assembly_fragment:04418"
misc_feature 162230..164443
/note="assembly_fragment:04428"

```

Query Match 70.4%; Score 1317.8; DB 80; Length 267581;

Best Local Similarity 91.9%; Pred. No. 0;

Matches 1355; Conservative 4; Mismatches 113; Indels 3; Gaps 3;

```

QY 277 cagagatccctgggagaaatgccgcgcgccatcttggtcatcgatgagcctcgccctg 336
|||
Db 165900 CACAGATCCCTGGGAGAAATGCCGCCGCCCATCTGGGTCATCGATGAGCCTGCCCTG 165841

QY 337 tgcctggctcccgctgtgagggaagacattagaaaaatgaattgatgttctctaaagg 396
|||
Db 165840 TGCCTGGTCCCGCTTGTGAGGGAAGACATTAGAAAATGAATGTATGTCTTCTTAAAGG 165781

QY 397 atgggcaggaaaaacagatcctctgttgatatattttgaacgggattacagatttgaaa 456
|||
Db 165780 ATGGGCAGGAAAACAGATCCTGTTGTGGATATTTATTGACGGGATTACAGATTGAAA 165721

QY 457 tgaagtccacaaagtgtgagcattaccatgagaggaacacagacgagaaaaatcttgaaggct 516
|||
Db 165720 TCAAGTCACAAAGTGAGCATTACCAATGAGAGGNAACACAGAGAAAATCTTGATGGCT 165661

QY 517 tcacaagacatgcacacaaacaaatggaatactgtgatgacatgagggcagcaagctggg 576
|||
Db 165660 TCACAAGACATGCAACAAACAAATGGAATACTGTGATGACATGAGCGACCAAGCTGGG 165601

QY 577 gagagataaacacgggcagaggtcaggaattctggccctgcctcctaaactgtcgtt 636
|||
Db 165600 GAGGAGATAACCCAGGCGAGGGTFCAGGATCTGGCCCTGCTGCCTAAACTGTGCGTT 165541

QY 637 catacccaatcatcttcatatttctaaccctcaaaacaaagctgtgttaatactgatct 696
|||
Db 165540 CATAACCAATCATTTTCATATTTCTAAACCTCAAAACAAAGCTGTTGTAATATCTGATCT 165481

QY 697 ctacgggttctcttgggcccacattctccatatatccagccacacactatcttataatt 756
|||
Db 165480 CTACGGTTCTCTTGGGCCCAACATTTCTCCATATATCCAGCCACACACTATTTTAAATAT 165421

QY 757 tagttccagatctatctactgtgaacctttctacacgtagaataacattactctttgtt 816
|||
Db 165420 TAGTTCACAGATCTGTACTGTGACCTTCTACACGTAGAAATACATTAATCTCATTTGTT 165361

QY 817 caaagacctctgtgtgtgctgctaataatgtagctgactgttttctcctaaaggagtgtct 876
|||
Db 165360 CAAAGACCTTCGTGTGTGCTGCTTAATATGTAGCTGACTGTTTTCCTAAGGAGTGTCT 165301

QY 877 gcccaggggatctgtgaacaggctgggaagcatctcaagatctttccagggtatactt 936
|||
Db 165300 GCCCCAGGGGATCTGTGAACAGGCTGGGAAGCATCTCAAGATCTTTCCAGGGTTATACCT 165241

QY 937 actagcacacagatgattacagagtggaattatctaatcaacatcatctcagtgctc 996
|||
Db 165240 ACTAGCACACAGATGATCATTTACGGAGTGAATATATCTAATCAACATCATCTCAGTGTC 165181

```

```
QY 997 ttgtcccaactgaattccattcccaacttttggccattctcccaactcctcaaatgtc 1056
Db 165180 TTTGCCCATACTGAAATTCATTCCCACTTTGTGCCATCTCAAGACCTCAAAATGTC 165121
QY 1057 attccattaatcacagattaaacttttttttttttttttttttttttttttttttt 1116
Db 165120 ATTCCATTAAATACAGAGTTAACTTTTCTTTTAACTGGAAGAAATTCATGTTTACAT 165061
QY 1117 gcagctatgggaatttaattacatatatttggttccagtgcaaatgactaaagcccttt 1176
Db 165060 GCAGCTATGGGAATTAATACATATTTGTTTCCAGTGCAGAAAGATGACTAAAGTCCCTT 165001
QY 1177 atccctccccctttgttgattttttccagttataaagttaaaatgcttagccttgact 1236
Db 165000 ATCCCTCCCCCTTTGTTGATTTTCTTTTCCAGTATAAAGTTAAATGCTTAGCCTTGACT 164941
QY 1237 gaggctgtatacag-cacagcctctcccccatccctccagccttatctgtcatcacatca 1295
Db 164940 GAGGCTGTATACAGCCACAGCCTCTCCCATCCCTCCAGCCTTATCTGTCAATCACCATCA 164881
QY 1296 accctcccatnysacctaaacaaactaaacttaacttgtaattccttgaaacatgtcagpnca 1355
Db 164880 ACCCTCCCATG-CACCTAAACAAATCTAATCTGTAATTCCTTGAAACATGTCAGG-CAT 164823
QY 1356 acattttccctctgcctgagaagctcttcttctctcttaantctagaatgatgataag 1415
Db 164822 ACATTATTCTTCTGCTGAGAAGCTCTTCCCTTGCTCTTAAATCTAGAATGATGATAAG 164763
QY 1416 ttttgaataagtgtactatttacttctaatgcaaaagagggacacatatagattcatcat 1475
Db 164762 TTTTGAATAAGTTGACTACTTCTTACTTTCATGCAAAAGAGGACACATATGAGATTCTATCAT 164703
QY 1476 cacatgacacagcaataactaaagttaattgtatttaagagtttagataaataatgtg 1535
Db 164702 CACATGACACAGCAAACTAACTAAAGTGTAATTTGATTAAGAGTTTAGATAAATATATG 164643
QY 1536 aaatgaagakccacagaggggaatttttatggggcagcttttgaagcctgggatgtgaag 1595
Db 164642 AAATGCAAGAGCCACAGAGGAATCTTTATGGGCACGCTTTGTAAGCCTGGGATGTGAAG 164583
QY 1596 maaagcagggaacctatagattattatataataataacttctctctctctctctctacac 1655
Db 164582 CAAAGCGAGGGAACCTCATGATCTTATATATATATTTNNNNNNNNNNNNNNNNNNNN 164523
QY 1656 aataccaacaagctttcacagaattcatgcagtgcaaaatcccaaggtaacctttat 1715
Db 164522 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 164463
QY 1716 ccatttcagtgtagtgcgcttagaattttggca 1750
Db 164462 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 164428

RESULT 6
LOCUS AX018075 359 bp DNA PAT 07-SEP-2000
DEFINITION Sequence 23 from Patent WO9946374.
ACCESSION AX018075
VERSION AX018075.1 GI:10042526
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 359)
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Pillarsky,C.
TITLE Human nucleic acid sequences from prostate tumour tissue
JOURNAL Patent: WO 9946374-A 23 16-SEP-1999.
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILLARSKY CHRISTIAN (DE)
FEATURES
Location/Qualifiers
```

```
source 1. .359
/db_xref="taxon:9606"
BASE COUNT 121 a 75 c 94 g 69 t
ORIGIN
Query Match 17.6%; Score 330; DB 9; Length 359;
Best Local Similarity 100.0%; Pred. No. 4.6e-71;
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 agaaactggcatcagaaaaacagaggagatttggctggctgcagggagagaccag 60
Db 30 AGAAGCTGGCATCAGAAAAACAGAGGGGAGATTGTTGTGGCTGCGAGCGGAGACCCAG 89
QY 61 gaaagctgcagtggtggaagacctgatgatacacagagtgagaataagaagcgtct 120
Db 90 GAAGATCTGCATGGTGGGAAGACCTGATGATACAGAGTGAGAATAAGAANGCTGCT 149
QY 121 gactttaccatctgagggcacacatctgctgaatggagataatatacatcactagaac 180
Db 150 GACTTTACCATCTGAGGCCACACATCTGCTGAATGGAGATAATTAAACATCACTAGAAC 209
QY 181 agcaagatgacaataatagtctaaagttagtgacatgtttttgacacatttccagcccttt 240
Db 210 AGCAAGATGACAATAATAGTCTAAAGTAGTGACATGTTTTCACATTTCCAGCCCTTT 269
QY 241 aaatattccacacacagaaagacaaaggaagacacagagatccctgggagaaatgcc 300
Db 270 AAATATCCACACACAGAGAGCAAAAGGAAGACACAGAGATCCCTGGGAGAAATGCC 329
QY 301 ggccgccatcttgggtcatcgtatgagcctc 330
Db 330 GCCCGCATCTGGGTCTCGATGAGCCTC 359

RESULT 7
LOCUS AL161625 143675 bp DNA HTG 20-JAN-2001
DEFINITION Homo sapiens chromosome 9 clone RP11-146P9, *** SEQUENCING IN
PROGRESS ***, 10 unordered pieces.
ACCESSION AL161625
VERSION AL161625.6 GI:9863607
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 143675)
AUTHORS Plumb,B.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT On Aug 21, 2000 this sequence version replaced gi:8894260.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bal46p9
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 138647 bases at least Q40
Consensus quality: 140619 bases at least Q30
Consensus quality: 141706 bases at least Q20
Insert size: 142775; sum-of-contigs
Insert size: 147523; 5.9% error; agarose-fp
Quality coverage: 4.29x in Q20 bases; sum-of-contigs Quality
coverage: 4.22x in Q20 bases; agarose-fp
```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 12163: contig of 12163 bp in length
* 12164 12263: gap of 100 bp
* 12264 25269: contig of 13006 bp in length
* 25270 25369: gap of 100 bp
* 25370 57709: contig of 32340 bp in length
* 57710 57809: gap of 100 bp
* 57810 66792: contig of 8983 bp in length
* 66793 68892: gap of 100 bp
* 68893 102772: contig of 35880 bp in length
* 102773 102872: gap of 100 bp
* 102873 106863: contig of 3991 bp in length
* 106864 106963: gap of 100 bp
* 106964 129804: contig of 22841 bp in length
* 129805 129904: gap of 100 bp
* 129905 132223: contig of 2319 bp in length
* 132224 133233: gap of 100 bp
* 133234 135764: contig of 3441 bp in length
* 135765 135864: gap of 100 bp
* 135865 143675: contig of 7811 bp in length.

FEATURES

Source
1. .143675
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone_lib="RPC1-11.1"
/clone="RP11-146P9"
misc_feature
1. .12163
/note="assembly_fragment:00440
vector_side:left"
clone_end:T7
misc_feature
12264..25269
/note="assembly_fragment:00434
fragment_chain:1"
misc_feature
25370..57709
/note="assembly_fragment:00526
fragment_chain:1"
misc_feature
57810..66792
/note="assembly_fragment:00661
fragment_chain:1"
misc_feature
66893..102772
/note="assembly_fragment:00323"
102873..106863
/note="assembly_fragment:00860"
106964..129804
/note="assembly_fragment:00978"
129905..132223
/note="assembly_fragment:01068"
132324..135764
/note="assembly_fragment:01424"
135865..143675
/note="assembly_fragment:01413
clone_end:SP6
vector_side:right"
BASE COUNT 42329 a 30042 c 28964 g 41431 t 909 others
ORIGIN

Query Match 5.4%; Score 100.2; DB 79; Length 143675;
Best Local Similarity 97.1%; Pred. No. 5.4e-14;
Matches 102; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 aqaagctgcacagaaacagagggagatttggctgcagccagggagaccag 60
|||||
Db 136065 AGNAGCTGCATCAGAAAAACAGAGGGGAGATTTGTGTGCTGCACCCGAGGAGACCAG 136124

QY 61 gaagatctcatggtgggaagacctgatgatacagaggtgagaa 105
|||||
Db 136125 GAAGATCTCATGCTGGGAGGACCTGATGATACAGAGGTCTGTA 136169
|||||
RESULT 8
AL358573
LOCUS AL358573 267581 bp DNA HTG 15-APR-2001
DEFINITION Homo sapiens chromosome 9 clone RP11-133022, *** SEQUENCING IN
PROGRESS ***, 37 unordered pieces.
ACCESSION AL358573
VERSION AL358573.17 GI:13660951
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 267581)
Mclay,K.
Direct Submission
Submitted (14-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Apr 17, 2001 this sequence version replaced gi:13398774.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bal33022
----- Summary Statistics
Assembly program: XCAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 256699 bases at least Q40
Consensus quality: 259744 bases at least Q30
Consensus quality: 261407 bases at least Q20
Insert size: 263981; sum-of-contigs
Insert size: 135491; 19.3% error; agarose-fp
Quality coverage: 5.51x in Q20 bases; sum-of-contigs Quality
coverage: 12.23x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 25718: contig of 25718 bp in length
* 25719 25818: gap of 100 bp
* 25819 29567: contig of 3749 bp in length
* 29568 29667: gap of 100 bp
* 29668 40520: contig of 10853 bp in length
* 40521 40620: gap of 100 bp
* 40621 50002: contig of 9382 bp in length
* 50003 50102: gap of 100 bp
* 50103 54942: contig of 4840 bp in length
* 54943 55042: gap of 100 bp
* 55043 59067: contig of 4025 bp in length
* 59068 59167: gap of 100 bp
* 59168 72327: contig of 13160 bp in length
* 72328 72427: gap of 100 bp
* 72428 79396: contig of 6969 bp in length
* 79397 79496: gap of 100 bp
* 79497 82738: contig of 3242 bp in length
* 82739 82838: gap of 100 bp
* 82839 87543: contig of 4705 bp in length
* 87544 87643: gap of 100 bp
* 87644 89847: contig of 2204 bp in length

```
* 89848 89947: gap of 100 bp
* 89948 92376: contig of 2429 bp in length
* 92377 92476: gap of 100 bp
* 92477 99799: contig of 7323 bp in length
* 99800 99899: gap of 100 bp
* 99900 102117: contig of 2218 bp in length
* 102118 102217: gap of 100 bp
* 102218 109330: contig of 7113 bp in length
* 109331 109430: gap of 100 bp
* 109431 112187: contig of 2757 bp in length
* 112188 112287: gap of 100 bp
* 112288 116407: contig of 4120 bp in length
* 116408 116507: gap of 100 bp
* 116508 119386: contig of 2879 bp in length
* 119387 119486: gap of 100 bp
* 119487 121869: contig of 2383 bp in length
* 121870 121969: gap of 100 bp
* 121970 125346: contig of 3577 bp in length
* 125347 125646: gap of 100 bp
* 125647 128990: contig of 3344 bp in length
* 128991 129090: gap of 100 bp
* 129091 134055: contig of 4965 bp in length
* 134056 134155: gap of 100 bp
* 134156 138314: contig of 4159 bp in length
* 138315 138414: gap of 100 bp
* 138415 140612: contig of 2198 bp in length
* 140613 140712: gap of 100 bp
* 140713 143296: contig of 2584 bp in length
* 143297 143396: gap of 100 bp
* 143397 148113: contig of 4717 bp in length
* 148114 148213: gap of 100 bp
* 148214 151009: contig of 2796 bp in length
* 151010 151109: gap of 100 bp
* 151110 156616: contig of 5507 bp in length
* 156617 156716: gap of 100 bp
* 156717 158763: contig of 2047 bp in length
* 158764 158863: gap of 100 bp
* 158864 162129: contig of 3266 bp in length
* 162130 162229: gap of 100 bp
* 162230 164443: contig of 2214 bp in length
* 164444 164543: gap of 100 bp
* 164544 169800: contig of 5257 bp in length
* 169801 169900: gap of 100 bp
* 169901 251545: contig of 81645 bp in length
* 251546 251645: gap of 100 bp
* 251646 259029: contig of 7384 bp in length
* 259030 259129: gap of 100 bp
* 259130 262871: contig of 3742 bp in length
* 262872 262971: gap of 100 bp
* 262972 265209: contig of 2238 bp in length
* 265210 265309: gap of 100 bp
* 265310 267581: contig of 2272 bp in length.
FEATURES
    source
        1..267581
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="9"
            /clone="RP11-133022"
            /clone_lib="RPC1-11.1"
            1..25718
                /note="assembly_fragment:00645"
                fragment_chain:1
                    clone_end:77
                    vector_side:left
                    25819..29567
                        /note="assembly_fragment:01660"
                        fragment_chain:1
                            29668..40520
                                /note="assembly_fragment:03080"
                                fragment_chain:1
                                    40621..50002
                                        /note="assembly_fragment:00771"
                                        fragment_chain:1
```

```
misc_feature
    /note="assembly_fragment:02629"
    fragment_chain:1
        55043..59067
            /note="assembly_fragment:04565"
            fragment_chain:1
                59168..72327
                    /note="assembly_fragment:00223"
                    fragment_chain:1
                        72428..79396
                            /note="assembly_fragment:03318"
                            fragment_chain:1
                                79497..82738
                                    /note="assembly_fragment:00224"
                                    fragment_chain:2
                                        82839..87543
                                            /note="assembly_fragment:03050"
                                            fragment_chain:2
                                                87644..89847
                                                    /note="assembly_fragment:01948"
                                                    fragment_chain:3
                                                        89948..92376
                                                            /note="assembly_fragment:02191"
                                                            fragment_chain:3
                                                                92477..99799
                                                                    /note="assembly_fragment:00010"
                                                                    99900..102117
                                                                        /note="assembly_fragment:00672"
                                                                        102218..109330
                                                                            /note="assembly_fragment:01326"
                                                                            109431..112187
                                                                                /note="assembly_fragment:01498"
                                                                                112288..116407
                                                                                    /note="assembly_fragment:01607"
                                                                                    116508..119386
                                                                                        /note="assembly_fragment:01899"
                                                                                        119487..121869
                                                                                            /note="assembly_fragment:02068"
                                                                                            121970..125546
                                                                                                /note="assembly_fragment:02469"
                                                                                                125647..128990
                                                                                                    /note="assembly_fragment:02502"
                                                                                                    129091..134055
                                                                                                        /note="assembly_fragment:02669"
                                                                                                        134156..138314
                                                                                                            /note="assembly_fragment:02757"
                                                                                                            138415..140612
                                                                                                                /note="assembly_fragment:02776"
                                                                                                                140713..143296
                                                                                                                    /note="assembly_fragment:03127"
                                                                                                                    143397..148113
                                                                                                                        /note="assembly_fragment:03634"
                                                                                                                        148214..151009
                                                                                                                            /note="assembly_fragment:03898"
                                                                                                                            151110..156616
                                                                                                                                /note="assembly_fragment:04220"
                                                                                                                                156717..158763
                                                                                                                                    /note="assembly_fragment:04276"
                                                                                                                                    158864..162129
                                                                                                                                        /note="assembly_fragment:04418"
                                                                                                                                        162230..164443
                                                                                                                                            /note="assembly_fragment:04428"
```

```
Query Match          5.4%; Score 100.2; DB 80; Length 267581;
Best Local Similarity 97.1%; Pred. No. 5.8e-14;
Matches 102; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 aqaagctgcatcagaaacagagggagattgtgtggtcagccagggagaccag 60
|||||
Db 253275 AGNAGCTGCATCAGAAAAACAGAGGGGAGATTGTGTGTCACCCGAGACCAG 253334
|||||

Qy 61 gaagatctgctggtgggaaggacctgatgatcacagaggtgagaa 105
|||||
```


misc_feature /note="assembly_fragment"
19424..24065
misc_feature /note="assembly_fragment"
24166..31658
misc_feature /note="assembly_fragment"
31759..38737
misc_feature /note="assembly_fragment"
38838..46131
misc_feature /note="assembly_fragment"
46232..52344
misc_feature /note="assembly_fragment"
52445..59763
misc_feature /note="assembly_fragment"
59864..66816
misc_feature /note="assembly_fragment"
66917..74010
misc_feature /note="assembly_fragment"
clone_end:T7
vector_side:left
74111..80599
misc_feature /note="assembly_fragment"
clone_end:SP6
vector_side:left
80700..89507
misc_feature /note="assembly_fragment"
89608..100337
misc_feature /note="assembly_fragment"
100438..114298
misc_feature /note="assembly_fragment"
114399..128530
misc_feature /note="assembly_fragment"
128631..142917
misc_feature /note="assembly_fragment"
143018..156243
misc_feature /note="assembly_fragment"
156344..172298
misc_feature /note="assembly_fragment"
BASE COUNT 51330 a 34914 c 34487 g 49065 t 2502 others
ORIGIN

Query Match 5.3%; Score 98.6; DB 61; Length 172298;
Best Local Similarity 96.2%; Pred. No. 1.4e-13;
Matches 101; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 agaactgcatcagaaaaacagagggagatttctgtgctgcagccgagggagaccag 60
Db 16539 AGAAGCTGGCATCAGAAAAACAGAGGGGAGATTCTGTGGCTGCAGCCGAGGGAGACCAG 16480
Qy 61 gaagatctgcattgtgggaagaccctgatgatacagagtgagaa 105
Db 16479 GAAGATTTCATGTTGGGAGGACCTGATGATACAGAGTCTGTA 16435

RESULT 10
AF279290
LOCUS AF279290 580 bp DNA PRI 04-DEC-2000
DEFINITION Homo sapiens prostate-cancer-specific DD3 protein gene, promoter region and exon 1.
ACCESSION AF279290
VERSION AF279290.1 GI:11528086
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 580)
AUTHORS Verhaegh,G.W., van Bokhoven,A., Smit,F., Schalken,J.A. and Bussemakers,M.J.G.
TITLE Isolation and Characterization of the Promoter of the Human Prostate Cancer-specific DD3 Gene
J. Biol. Chem. 275 (48), 37496-37503 (2000)
PUBMED 10982808

REFERENCE 2 (bases 1 to 580)
AUTHORS Verhaegh,G.W., van Bokhoven,A., Smit,F., Schalken,J.A. and Bussemakers,M.J.G.
TITLE Direct Submission
JOURNAL Submitted (16-JUN-2000) Urology Research Laboratory, University Medical Center Nijmegen, Geert Grooteplein Zuid 10, Nijmegen 6525 GA, The Netherlands
FEATURES
source Location/Qualifiers
1..580
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/map="9q21-q22"
1..460
promoter 260..288
protein_bind /bound_moiety="high mobility group I Y protein"
461..580
mRNA /product="prostate-cancer-specific DD3 protein"
461..580
exon /number=1
BASE COUNT 190 a 97 c 153 g 140 t
ORIGIN
Query Match 5.2%; Score 98; DB 89; Length 580;
Best Local Similarity 100.0%; Pred. No. 1.1e-13;
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 agaactgcatcagaaaaacagagggagatttctgtgctgcagccgagggagaccag 60
Db 483 AGAAGCTGGCATCAGAAAAACAGAGGGGAGATTCTGTGGCTGCAGCCGAGGGAGACCAG 542
Qy 61 gaagatctgcattgtgggaagaccctgatgatacagag 98
Db 543 GAAGATCTGCATGTTGGGAGGACCTGATGATACAGAG 580
RESULT 11
I66494
LOCUS I66494 7218 bp DNA PAT 28-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION I66494
VERSION I66494.1 GI:2724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dörner,F., Scheiflinger,F. and Falkner,F.Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES Location/Qualifiers
source 1..7218
/organism="unknown"
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN
Query Match 3.2%; Score 59.8; DB 10; Length 7218;
Best Local Similarity 5.1%; Pred. NO. 0.00039;
Matches 20; Conservative 219; Mismatches 155; Indels 0; Gaps 0;
Qy 1013 ttcatctccacttttggccattctcaagacacctcaaatgtcattccattaatcac 1072
Db 1076 YY 1135
Qy 1073 aggattaaactttttttaacctggaagaattcaatgttacatgcagctatgggaatt 1132
Db 1136 YY 1195
Qy 1133 aattacatatattttttccagtcgaaagatgactaaagtcctttatccctccctttgt 1192
Db 1196 YY 1255

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 04:11:46 ; Search time 6114.61 Seconds
(without alignments)
1267.675 Million cell updates/sec

Title: US-09-402-713A-4
Perfect score: 820
Sequence: 1 agaagctggcatcagaaaaa.....cattactcattgttcaaa 820

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues
Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.


```

DEFINITION RC5-FT0193-201100-012-D06 FT0193 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF858286
VERSION BF858286.1 GI:12246030
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 332)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC5&t2=RC5-FT0193-
201100-012-D06&t3=2000-11-20&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 332.
Location/Qualifiers
1. 332
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="FT0193"
/dev_stage="Adult"
/note="Organ: prostate_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 90 a 84 c 64 g 94 t
ORIGIN
Query Match 35.5%; Score 290.8; DB 170; Length 332;
Best Local Similarity 99.3%; Pred. No. 3.2e-71;
Matches 292; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 527 tgcacaaacaaatggaatactgtgtgacatgagcgagcgcaagctggggaggagataa 586
|||||
Db 12 TGCACAAACAAATGAATGACTGTGTGATGACATGAGCGAGCCCAAGCTGGGAGGAGATAA 71
|||||

QY 587 ccacggggcagagggtcaggattctgctgctgcctaaactgtcgttcataaccaa 646
|||||
Db 72 CCACGGGGCAGAGGTCAGAGATTCTGGCCCTCTGCTAACTAGTCGGTTCAATACCAA 131
|||||

QY 647 tcaattcattatttcaacctcaaaacaaagctgtgtgtaatatctgtactctacggttcc 706
|||||
Db 132 TCATTTTCATATTTCTAACCTCAAAACAAAGCTGTGTGTAATATCTGATCTCTAGGGTTCC 191
|||||

QY 707 ttctggggcccaacattctccatatatccagccacactcatttttaattattagttccag 766
|||||
Db 192 TTTCTGGGCCCAACATTTCTCCATATATCCAGCCACACTCATTTTAAATATTTAGTCCAG 251
|||||

QY 767 atctgtactgtgacctttctacactgtagaataaacattactctatttcttcaaa 820
|||||
Db 252 ANCTGTACTGTGACCTTCTACGCTGTAGTAACATTAACATTAATCTATTGTTTCANA 305
|||||

```

```

RESULT 3
BF373619
LOCUS BF373619
DEFINITION MR0-FT0175-310800-106-h09 FT0175 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF373619
VERSION BF373619.1 GI:11335644
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 290)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR0&t2=MR0-FT0175-
310800-106-h09&t3=2000-08-31&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 290.
Location/Qualifiers
1. 290
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="FT0175"
/dev_stage="Adult"
/note="Organ: prostate_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 78 a 73 c 53 g 86 t
ORIGIN
Query Match 29.8%; Score 244.4; DB 147; Length 290;
Best Local Similarity 99.6%; Pred. No. 3.7e-58;
Matches 245; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 575 gggaggagataaccacggggcgagggtcaggattctggtccctgctgcctaactgtgcg 634
|||||
Db 21 GGGAGGAGATAACACGGGGCAGAGGGTCAGGATTCTTGCCCTGTGCTCAACTGTGGC 80
|||||

QY 635 ttcataaccaaattcatttcatttcctaacctcaaaacaaagctgtgtataatctgat 694
|||||
Db 81 TTCATAACCAATTCATTCATATTTCTAACCTCAAAACAAAGCTGTGTGTAATATCTGAT 140
|||||

QY 695 ctctacggttctcttggggcccaacattctccatatatccagccacactcttttaata 754
|||||
Db 141 CTCTACGGTCTCTTCTGGGCCCAACATTTCTCCATATATCCAGCCACACTCATTTTAAATA 200
|||||

QY 755 tttagttccagatctgtactgtgacctttctacactgtagaataaacattactctatttg 814
|||||

```

Db 201 TTAGTCCGATCTGTACTGTGACCTTTCTACACTGTAGAATAACATTAACATTTGCG 260

Qy 815 ttcaaa 820

Db 261 TTCAAA 266

RESULT 4

LOCUS BF373581 262 bp mRNA EST 24-NOV-2000

DEFINITION MRO-FT0175-210800-101-d05 FT0175 Homo sapiens cDNA, mRNA sequence.

ACCESSION BF373581

VERSION BF373581.1 GI:11335606

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 282)

AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

COMMENT Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR04t2=MR0-FT0175-210800-101-d05&t3=2000-08-21&t4=1)

Qy 210800-101-d05&t3=2000-08-21&t4=1

Seg primer: puc 18 forward

High quality sequence start: 35

High quality sequence stop: 282.

FEATURES

source

1. 282

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="FT0175"

/dev_stage="Adult"

/note="Organ: prostate_tumor; Vector: puc18; Site:1: SmaI; Site:2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 85 a 51 c 70 g 76 t

ORIGIN

Query Match 27.5%; Score 225.6; DB 147; Length 282;

Best Local Similarity 95.7%; Pred. No. 7.5e-53;

Matches 243; Conservative 0; Mismatches 9; Indels 2; Gaps 1;

Qy 568 caagctggggaggagataaccacggggcagagggtcaggattctggccctgctgcctaaa 627

Db 282 CAAGCTGGGAGGAGATAAACACGGGAGAGGGTCAGGATTCTGCGCCCTGCTGCTRAA 223

Qy 628 ctgtgcgttcataacaaatcatttcatttcataaccctcaaaacaaagctgtgtaat 687

Db 222 CTGTGCGTTTCAACAAATCATTTTCATATTTCTAACCCCTCAAAACAAAGCTGTGTAAT 163

Qy 688 atctgatctacaggttcctctggggcccaacattctccatatatccagccacctcatt 747

Db 162 ATCTGATCTCTACGGTTCCTTCTGGGCCCAACATTTCTCATATATCCAGCCACACTCATT 103

Qy 748 tttataattagttcccgagatctgtactgtgaccttctc--tacactgtagaataacatta 805

Db 102 TTTAATATTAGTTCCAGATCTGTACTGTGACCTTTCTACATCTGTAGAATAACATTA 43

Qy 806 ctcatttggitcaa 819

Db 42 CTCATTGTTCAAA 29

RESULT 5

LOCUS BF58371 226 bp mRNA EST 16-JAN-2001

DEFINITION RCS-FT0193-211100-012-El1 FT0193 Homo sapiens cDNA, mRNA sequence.

ACCESSION BF58371

VERSION BF58371.1 GI:12246115

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 226)

AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

COMMENT Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC5&t2=RC5-FT0193-211100-012-El1&t3=2000-11-21&t4=1)

Seg primer: puc 18 forward

High quality sequence stop: 226.

FEATURES

source

1. 226

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="FT0193"

/dev_stage="Adult"

/note="Organ: prostate_tumor; Vector: puc18; Site:1: SmaI; Site:2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 79 a 29 c 59 g 59 t

ORIGIN

Query Match 24.3%; Score 199; DB 170; Length 226;

Best Local Similarity 100.0%; Pred. No. 2.2e-45;

Matches 199; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 622 cctaaactgtgcgttcataacaaatcatttcatttcataaccctcaaaacaaagctgt 681

Db 226 CCTAAACTGTGCGTTTCATAAACCAATCATTTTCATATTTCTAACCCCTCAAAACAAAGCTGT 167

```

QY 682 tgaataatgatctctacaggttcctctcttggtggcccaacattccatataatccagcaca 741
|||||
Db 166 TGAATATCTGATCTACGGTTCCTCTGGGCCCAACATTCCTCATATATCCAGGCACA 107
|||||
QY 742 ctcatitttaattagttccagatctgactgtgacctttctacactgtagaataac 801
|||||
Db 106 CTCAATTTTAATATTTAGTTCCAGATCTGACTGTGACCTTCTTACACTGTAGATAAC 47
|||||
QY 802 attactcattttgttcaaa 820
|||||
Db 46 ATTACTATTTTGTCAAA 28
|||||

RESULT 6
BF373406 167 bp mRNA EST 24-NOV-2000
LOCUS IL2-FT0159-070800-120-H01 FT0159 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF373406
VERSION BF373406.1 GI:11335431
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 167)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=IL2&t2=IL2-FT0159-
070800-120-H01&t3=2000-08-07&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 167.
FEATURES
source
1..167
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="FT0159"
/dev_stage="Adult"
/note="Organ: prostate_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 47 a 38 c 49 g 33 t
ORIGIN
Query Match 20.4%; Score 167; DB 147; Length 167;
Best Local Similarity 100.0%; Pred. No. 2.2e-36;
Matches 167; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 254 cacagaagcacaaggaagcacagagatccctggagaaatgccggccgacctgtg 313
|||||

```

```

Db 1 CACAGGACGACAAAAGGAGGACAGAGATCCCTGGAGAAATGCCGGCGCCATCTTG 60
QY 314 ggtcatcgatgagcctgcgcctgtgctccgtgctcgttgagggaagacattagaaaa 373
|||||
Db 61 GGTCACTGATGAGCCTCGCCCTGTGCTGCTCCGCTTGAGGGAAGGACATTAGAAAA 120
|||||
QY 374 tgaattgatgtgtctcttaagagatggcagggaacacagatcctgtt 420
|||||
Db 121 TGAATTGATGTGTCTCTTAAGGATGGCGAGGAAAAACAGATCCTGTT 167
|||||

RESULT 7
AQ206972/c 394 bp DNA GSS 17-SEP-1998
LOCUS HS_3238_B1_G11_MR CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate=3238 Col=21 Row=N, DNA sequence.
ACCESSION AQ206972
VERSION AQ206972.1 GI:3617542
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 394)
Mahairs,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairs GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence tagged Connector
Plate: 3238 row: N column: 21
Class: BAC ends
High quality sequence stop: 394.
FEATURES
source
1..394
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Plate=3238 Col=21 Row=N"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 134 a 62 c 93 g 103 t 2 others
ORIGIN
Query Match 13.7%; Score 112.4; DB 225; Length 394;
Best Local Similarity 97.4%; Pred. No. 7.1e-21;
Matches 113; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 705 cctctggcccaacattccatataatccagccacacatttttaattagttccc 764
|||||
Db 394 CCGCTGGGCNCACATTCCTCATATATCCAGCCACACTATTTTAATATTAGTTCCC 335
|||||
QY 765 agatctgtactgtgaccttctcacatgtagaataacattactctattgttcaaa 820
|||||
Db 334 AGATCTGTACTGTGACCTTCTTACACTGTAGATAACATTACTATTTGTTCAA 279
|||||

RESULT 8
AI557495 657 bp mRNA EST 09-AUG-1999
LOCUS PT2.1.7.H12.r tumor2 Homo sapiens cDNA 3', mRNA sequence.
ACCESSION AI557495

```



```
BASE COUNT      333 a      162 c      148 g      177 t      281 others
ORIGIN

Query Match      5.0%; Score 41; DB 219; Length 1101;
Best Local Similarity 21.9%; Pred. No. 1.3;
Matches 75; Conservative 113; Mismatches 155; Indels 0; Gaps 0;

QY 352 gtgaggaaggacatagaaatgaattgattgtctctaaagatggcagagaaaca 411
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 713 GKAGARARGCGGGRGAWRAAAKTKKTKTKRRAAGRRARRAWAGAAAAAATA 772
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 412 gattcgtgttgatatttttgaacgggattacagagattgaaatgaagtcacaaagt 471
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 773 DAATKTGKTWAKAAWRTAAWKDKWKATDAAAATAAARTDWTAKAKDKTKWGAATA 832
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 472 agcattaccatgagaggaagaaacagacagagaaatttggatgcttcacaaagcatg 531
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 833 WGAGAARGRWGRKGDGTTRARRAGAGDGDWAKAAWAAWAAATTAATTAADDDR 892
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 532 caacaaatggaatctgtgacatgagcagcagcaagctgggagagataaacacg 591
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 893 AKAWDRRAAARDKDKGRRAGTGWRRRARARTKAKAKRGARAGAKARRAKGDA 952
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 592 gggcagaggtcaggattctgcccctgctgcctaaactgtgcttcataaccacaaatcatt 651
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 953 WKDGWKKWKDKDAWAAKAKGTGDKWRKARKAKTGWKARADKDRWRAAADTRWRTKAW 1012
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 652 tcatattctaacccctcaaaacaaagctgtgtgtaatatctgat 694
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1013 WTDWRWATATKDKTKDKAKKKGKAWKTKTTRTDRAAKAKAKDDT 1055
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 15
AI427279
LOCUS      AI427279      481 bp      mRNA      EST      15-MAR-2000
DEFINITION mc86g07.y1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
ACCESSION  AI427279
VERSION    AI427279.1 GI:4273205
KEYWORDS  EST.
SOURCE    house mouse.
ORGANISM  Mus musculus
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS   1 (bases 1 to 481)
          Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
          Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
          ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
          ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
          Waterston,R. and Wilson,R.
          The WashU-NCI Mouse EST Project 1999
          Unpublished (1999)
          Contact: Marra M/WashU-NCI Mouse EST Project 1999
          Washington University School of Medicine
          4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
          Tel: 314 286 1800
          Fax: 314 286 1810
          Email: mouseest@watson.wustl.edu
          This clone is available royalty-free through LNL; contact the
          IMAGE Consortium (info@image.llnl.gov) for further information.
          This read is a RESEQUENCE of a previously sequenced mouse clone
          This read has been verified (found to hit its original self in the
          correct orientation)
          MGI:227252
          Seq primer: -40RP from Gibco
          High quality sequence stop: 454
          POLYA=No.
          Location/Qualifiers
            1..481
              /organism="Mus musculus"
              /strain="C57BL/6J"
```

```
FEATURES
Source
/db_xref="taxon:10090"
/clone="IMAGE:355452"
/clone_lib="Soares mouse embryo NbME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1; Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTCACCAATCTGAATGGACGGCCGGGAAATTTTTTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT7T3 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M. Fatima Bonaldo."
```

```
BASE COUNT      122 a      131 c      119 g      109 t
ORIGIN

Query Match      4.9%; Score 40.4; DB 20; Length 481;
Best Local Similarity 61.3%; Pred. No. 1.5;
Matches 65; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 447 agatttgaatgaagtccacaaagtgcattaccatgagagagagagagagagaaat 506
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 AAATTAGGAAAGAGGAGGAAAAAGAAATTTAAAAAAGAAAAAGAAAGATGAC 60
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 507 cttgatggcttcacagagatgcacacaaatggaatactgtg 552
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 CTGTGATGGAAAAAATAATTTAAAAAAGAAAAAGATATATCTGTG 106
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

Search completed: July 31, 2001, 04:11:52
Job time: 10004 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 07:20:07 ; Search time 176.8 seconds
(without alignments)
878.027 Million cell updates/sec

Title: US-09-402-713A-4
Perfect score: 820
Sequence: 1 agaagctggcatcagaaaaa.....cattactcattgttccaag 820

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	57.4	7.0	7218	1	US-08-232-463-14
2	35.6	4.3	7218	1	US-08-232-463-14
3	34.2	4.2	2679	1	US-07-977-434-11
4	34.2	4.2	2679	1	US-08-458-819-11
5	34.2	4.2	2679	5	PCT-US91-07035-11
C 6	33.8	4.1	246240	2	US-08-724-394A-20
C 7	33.8	4.1	246240	2	US-08-724-394A-21
C 8	33.8	4.1	246240	2	US-08-724-394A-22
C 9	32.4	4.0	713	3	US-08-532-896-26
10	31	3.8	2737	3	US-08-389-564B-5
11	31	3.8	2737	3	US-08-466-047B-5
12	30.6	3.7	466	3	US-08-899-437-21
13	30.6	3.7	466	4	US-09-126-121-21
14	30.6	3.7	2091	3	US-08-899-437-22
15	30.6	3.7	2091	4	US-09-126-121-22
16	30.6	3.7	2502	3	US-08-899-437-5
17	30.6	3.7	2502	4	US-09-126-121-5
18	30.6	3.7	4758	3	US-09-191-647-1
C 19	30.6	3.7	43795	3	US-08-742-185-101
20	30.4	3.7	877	3	US-09-129-888-1
C 21	30.4	3.7	2233	1	US-08-129-129-2
C 22	30.2	3.7	29604	3	US-08-781-891-207
C 23	29.8	3.6	9997	1	US-08-246-982A-15
C 24	29.8	3.6	9997	1	US-08-453-265-15
C 25	29.8	3.6	10103	2	US-08-457-273B-7
C 26	29.8	3.6	10607	1	US-08-078-090-3
C 27	29.8	3.6	35100	1	US-08-306-691B-19

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232.463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935.313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REFERENCE NUMBER: 29,768
; REGISTRATION NUMBER: 30472/114 INMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base-pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Sequence 19, Appl
Sequence 7, Appl
Sequence 18, Appl
Sequence 18, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 4, Appl
Sequence 1, Appl
Sequence 8, Appl
Sequence 1, Appl
Sequence 11, Appl
Sequence 11, Appl
Sequence 4, Appl
Sequence 11, Appl
Sequence 8, Appl
Sequence 14, Appl


```

; FILING DATE: 17-JUN-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 899,241
; FILING DATE: 22-AUG-1986
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 746,121
; FILING DATE: 15-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US90/07641
; FILING DATE: 21-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 585,471
; FILING DATE: 20-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 455,611
; FILING DATE: 22-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 609,157
; FILING DATE: 02-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 557,517
; FILING DATE: 24-JUL-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Luann Cserr
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: Case No. 5466591 8753
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 814-2972
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2679 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Thermosipho africanus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2676
; US-07-977-434-11

Query Match 4.2%; Score 34.2; DB 1; Length 2679;
Best Local Similarity 54.3%; Pred. No. 0.75;
Matches 69; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 102 agaaataagaaggctgctgaccttaccatctgaggccacacatctgctgaaatggagat 161
    ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2415 AGGAGAAAGAAATAGCTGTTAAACCTCCATTCAGGAAACAGCAGCTCATATAAAGAT 2474
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 162 aattaacatcactagaacagcaagatgacaataataatgctctaagtagtgacatgtttt 221
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2475 AGCTATGATTATATTCATAATAGATTGAAGAAGGAAAAATCTACGTTCAAAAATGATATT 2534
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 222 gcacatt 228
    ||| ||
DB 2535 GCAGGTT 2541

RESULT 4
US-08-458-819-11
; Sequence 11, Application US/08458819
; Patent No. 5795762
; GENERAL INFORMATION:
; APPLICANT: Gelfand, David H.
; APPLICANT: Abranson, Richard D.
; TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF
; THERMOSTABLE DNA POLYMERASES
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
;
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; ZIP: 07110-1199
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: 7
; SOFTWARE: WordPerfect 2.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,819
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/977,434
; FILING DATE: 23-FEB-1993
; APPLICATION NUMBER: US 590,490
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 590,466
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 590,213
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 523,394
; FILING DATE: 15-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 143,441
; FILING DATE: 12-JAN-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 063,509
; FILING DATE: 17-JUN-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 899,241
; FILING DATE: 22-AUG-1986
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 746,121
; FILING DATE: 15-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US90/07641
; FILING DATE: 21-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 585,471
; FILING DATE: 20-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 455,611
; FILING DATE: 22-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 609,157
; FILING DATE: 02-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 557,517
; FILING DATE: 24-JUL-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Luann Cserr
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: Case No. 5795762 8753
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 814-2972
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2679 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Thermosipho africanus
; FEATURE:
```


;
; TITLE OF INVENTION: Sequences and Antibodies Thereto
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND AND TOWNSEND AND CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONTIG"
; US-08-724-394A-20

Query Match 4.1%; Score 33.8; DB 2; Length 246240;
Best Local Similarity 44.2%; Pred. No. 12;
Matches 140; Conservative 0; Mismatches 177; Indels 0; Gaps 0;
QY 369 gaaatgaattgatgtgttccttaagagatggcgaggaacacagatccctgtgttgatat 428
DB 36625 GGAACCTCAATTAAAGGAAGTCTGAAGCTGATAAGCCAGAGAGGGAAGGCTCTCATTTCAATT 36566
QY 429 ttatttgaacggattacagatttgaatgaagtcacaaagtgcacaaatggaatgagag 488
DB 36565 TTATAAGGGTTGCGTCACACTAGGAAGATCCAATAGCAACACAGTCTCAAAATTAATGA 36506
QY 489 gaaacagacgagaaattcttgatgcttcacaaagacatgcacaaacaaatgggaatc 548
DB 36505 TTACAATAGGACACAAATTCGAAGCTCGGAGCCCAAGCAGAAATGGATTAGGGAAGAC 36446
QY 549 tbtgatgacatgagcgacgaagctgggagagagataaccacggggcagaggtcaggat 608
DB 36445 ATGGATGATATGAACAGGAGGAGGGGTACAAGGCAGCTTCTCTGGGAAGTTGCCAGGGC 36386
QY 609 tctggccctgctcctaaactgtgcttcataacaaatcatttcatttcaaccctc 668
DB 36385 AGTCACAGTTCACATTCATTAGGCTGTGGGCACCAATGCATATGGAAAAATCTAGCTGAC 36326
QY 669 aaaaacaaagctgttga 685
DB 36325 TTAACCTGAACCTCTGAA 36309

RESULT 7
US-08-724-394A-21/c
; Sequence 21, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1

;
; TITLE OF INVENTION: Sequences and Antibodies Thereto
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND AND TOWNSEND AND CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONTIG"
; US-08-724-394A-21

Query Match 4.1%; Score 33.8; DB 2; Length 246240;
Best Local Similarity 44.2%; Pred. No. 12;
Matches 140; Conservative 0; Mismatches 177; Indels 0; Gaps 0;
QY 369 gaaatgaattgatgtgttccttaagagatggcgaggaacacagatccctgtgttgatat 428
DB 36625 GGAACCTCAATTAAAGGAAGTCTGAAGCTGATAAGCCAGAGAGGGAAGGCTCTCATTTCAATT 36566
QY 429 ttatttgaacggattacagatttgaatgaagtcacaaagtgcacaaatggaatgagag 488
DB 36565 TTATAAGGGTTGCGTCACACTAGGAAGATCCAATAGCAACACAGTCTCAAAATTAATGA 36506
QY 489 gaaacagacgagaaattcttgatgcttcacaaagacatgcacaaacaaatgggaatc 548
DB 36505 TTACAATAGGACACAAATTCGAAGCTCGGAGCCCAAGCAGAAATGGATTAGGGAAGAC 36446
QY 549 tbtgatgacatgagcgacgaagctgggagagagataaccacggggcagaggtcaggat 608
DB 36445 ATGGATGATATGAACAGGAGGAGGGGTACAAGGCAGCTTCTCTGGGAAGTTGCCAGGGC 36386
QY 609 tctggccctgctcctaaactgtgcttcataacaaatcatttcatttcaaccctc 668
DB 36385 AGTCACAGTTCACATTCATTAGGCTGTGGGCACCAATGCATATGGAAAAATCTAGCTGAC 36326
QY 669 aaaaacaaagctgttga 685
DB 36325 TTAACCTGAACCTCTGAA 36309
RESULT 8
US-08-724-394A-22/c
; Sequence 22, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:

RESULT 10
US-08-389-564B-5
: Sequence 5: Application us/08389564B

```

; Patent No. 5989870
; GENERAL INFORMATION:
; APPLICANT: Nakari, Tiina H.
; APPLICANT: Onnela, Maija-Leena
; APPLICANT: Ilm n, Marja H.
; APPLICANT: Penttil , Merja E.
; TITLE OF INVENTION: A METHOD FOR CLONING ACTIVE PROMOTERS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US 08/389,564B
; FILING DATE: 16-FEB-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/932,485
; FILING DATE: 19-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/496,155
; FILING DATE: 19-MAR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/044,077
; FILING DATE: 29-APR-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 86 10600
; FILING DATE: 30-APR-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, GRANT E.
; REGISTRATION NUMBER: 41,264
; REFERENCE/DOCKET NUMBER: 1716.008000G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2737 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-389-564B-5

```

```

Query Match      3.8%; Score 31; DB 2; Length 2737;
Best Local Similarity 56.3%; Pred. No. 7.6;
Matches 58; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 50 agggagaccagaagatctgcatgtggaagacctaatacagaggtgagaataa 109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2056 AGAGGCATCTGGAGACACAGTATGGGGGAGGAGGATGGACCATTAAGTGGCGGGTGATTA 2115

Qy 110 gaaaggctgctgactttaccatctgagggccacacacatctgctga 152
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2116 CCATGACTCGGACTCGGACCAAGTGGAGGGGAGGCGCATCTGCTCA 2158

RESULT 11
US-08-466-047B-5
; Sequence 5, Application US/08466047B
; Patent No. 6011147
; GENERAL INFORMATION:
; APPLICANT: Nakari, Tiina H.
; APPLICANT: Onnela, Maija-Leena
; APPLICANT: Ilm n, Marja H.

```

```

; APPLICANT: Nevalainen, Kaisu Milja Helena
; APPLICANT: Penttil , Merja E.
; TITLE OF INVENTION: Fungal Promoters Active In The Presence
; TITLE OF INVENTION: Of Glucose
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US 08/466,047B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/389,564
; FILING DATE: 16-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/932,564
; FILING DATE: 19-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/496,155
; FILING DATE: 19-MAR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/044,077
; FILING DATE: 29-APR-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 86 10600
; FILING DATE: 30-APR-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, GRANT E.
; REGISTRATION NUMBER: 41,264
; REFERENCE/DOCKET NUMBER: 1716.008000H
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2737 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-466-047B-5

Query Match      3.8%; Score 31; DB 3; Length 2737;
Best Local Similarity 56.3%; Pred. No. 7.6;
Matches 58; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 50 agggagaccagaagatctgcatgtggaagacctaatacagaggtgagaataa 109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2056 AGAGGCATCTGGAGACACAGTATGGGGGAGGAGGATGGACCATTAAGTGGCGGGTGATTA 2115

Qy 110 gaaaggctgctgactttaccatctgagggccacacacatctgctga 152
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2116 CCATGACTCGGACTCGGACCAAGTGGAGGGGAGGCGCATCTGCTCA 2158

RESULT 12
US-08-899-437-21
; Sequence 21, Application US/08899437
; Patent No. 6121415
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
; TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
; TITLE OF INVENTION: Ligands and Uses Therefor

```

```

; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,437
; FILING DATE: 24-Jul-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Conley, Deirdre L.
; REGISTRATION NUMBER: 36,487
; REFERENCE/DOCKET NUMBER: P1084R1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-2066
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 466 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; FEATURE:
; NAME/KEY: EST Genbank entry H23651
; LOCATION: 1-466
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; US-08-899-437-21

Query Match 3.7%; Score 30.6; DB 3; Length 466;
Best Local Similarity 53.8%; Pred. No. 3.8;
Matches 63; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 450 ttgaaatgaagtcacaaagtgcattaccatgagagagaaacacagacgagaaaatctt 509
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159 TTCAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 218
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 510 gatgcttcacagacatcaacaaatgaatactgtgatgacatgagcgag 566
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 219 AAAAGTACAGTCTCAAGCATCCAGCACAAATGGCAAGTCAGAGAACTTGGTGAAG 275

RESULT 13
US-09-126-121-21
; Sequence 21, Application US/09126121
; Patent No. 6252051
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
; TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
; TITLE OF INVENTION: Ligands and Uses Therefor
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
```

```

; APPLICATION NUMBER: US/09/126,121
; FILING DATE: 30-Jul-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Conley, Deirdre L.
; REGISTRATION NUMBER: 36,487
; REFERENCE/DOCKET NUMBER: P1084R1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-2066
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 466 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; FEATURE:
; NAME/KEY: EST Genbank entry H23651
; LOCATION: 1-466
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; US-09-126-121-21

Query Match 3.7%; Score 30.6; DB 4; Length 466;
Best Local Similarity 53.8%; Pred. No. 3.8;
Matches 63; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 450 ttgaaatgaagtcacaaagtgcattaccatgagagagaaacacagacgagaaaatctt 509
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159 TTCAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 218
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 510 gatgcttcacagacatcaacaaatgaatactgtgatgacatgagcgag 566
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 219 AAAAGTACAGTCTCAAGCATCCAGCACAAATGGCAAGTCAGAGAACTTGGTGAAG 275

RESULT 14
US-08-899-437-22
; Sequence 22, Application US/08899437
; Patent No. 6121415
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
; TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
; TITLE OF INVENTION: Ligands and Uses Therefor
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,437
; FILING DATE: 24-Jul-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Conley, Deirdre L.
; REGISTRATION NUMBER: 36,487
; REFERENCE/DOCKET NUMBER: P1084R1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-2066
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2091 base pairs
; TYPE: Nucleic Acid
```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 07:26:09 ; Search time 478.87 Seconds
(without alignments)
1075.196 Million cell updates/sec

Title: US-09-402-713A-4
Perfect score: 820
Sequence: 1 agagctgcatcagaaaa.....cattactcattttgttcaaa 820

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues
Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601.*
1: /SID88/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	820	100.0	820	19 AAV62429	Prostate cancer an
2	820	100.0	1872	19 AAV62428	Prostate cancer an
3	812	99.0	812	21 AAA06690	Human immunogenic
4	725.8	88.5	3112	21 AAA06687	Human immunogenic
5	725.6	88.5	2229	21 AAA06688	Human immunogenic
6	725.6	88.5	2426	21 AAA06689	Human immunogenic
7	725.4	88.5	2037	19 AAV62427	Prostate cancer an
8	725.4	88.5	3582	19 AAV62430	Prostate cancer an
9	523.6	63.9	597	20 AAX37486	Human secreted pro
10	513.2	62.6	718	21 AAA06545	Human immunogenic
11	414.2	50.5	437	21 AAC05768	Human secreted pro

c 12	330	40.2	359	20	AAZ33445	Human prostate can
c 13	288.4	35.2	301	21	AAA06520	Human immunogenic
c 14	100	12.2	123	19	AAV33791	Prostate cancer an
c 15	76.4	9.3	936	22	AAF58252	Oligonucleotide D1
c 16	76.4	9.3	936	22	AAF58254	Oligonucleotide D1
c 17	76.4	9.3	936	22	AAF58257	Oligonucleotide D1
c 18	76.4	9.3	936	22	AAF58259	Oligonucleotide D2
c 19	76.4	9.3	936	22	AAF58262	Oligonucleotide D2
c 20	76.4	9.3	938	22	AAF58255	Oligonucleotide D1
c 21	71	8.7	936	22	AAF58252	Oligonucleotide D1
c 22	71	8.7	936	22	AAF58254	Oligonucleotide D1
c 23	71	8.7	936	22	AAF58257	Oligonucleotide D1
c 24	71	8.7	936	22	AAF58259	Oligonucleotide D2
c 25	71	8.7	936	22	AAF58262	Oligonucleotide D2
c 26	71	8.7	938	22	AAF58255	Oligonucleotide D1
c 27	39.8	4.9	244	22	AAF58238	Oligonucleotide D1
c 28	38.4	4.7	244	22	AAF58238	Oligonucleotide D1
c 29	36	4.4	254	21	AAC16115	Human secreted pro
c 30	35.8	4.4	67212	21	AAA08954	WFS1 variant genom
c 31	35.2	4.3	2147	21	AAZ60617	DNA encoding the p
c 32	34.6	4.2	982	20	AAX37417	Human secreted pro
c 33	34.2	4.2	1830	13	AAQ24334	Mutant thermostabl
c 34	34.2	4.2	2073	13	AAQ24333	Mutant thermostabl
c 35	34.2	4.2	2265	13	AAQ24332	Mutant thermostabl
c 36	34.2	4.2	2403	13	AAQ24331	Mutant thermostabl
c 37	34.2	4.2	2568	13	AAQ28937	Encodes Taf DNA po
c 38	34.2	4.2	2571	13	AAQ24330	Mutant thermostabl
c 39	34.2	4.2	2679	13	AAQ24329	Mutant thermostabl
c 40	34.2	4.2	2679	13	AAQ28936	Encodes Asp37 Taf
c 41	34.2	4.2	4286	13	AAQ23917	Taf DNA polymerase
c 42	34.2	4.2	122186	22	AAC89560	Human histone deac
c 43	33.8	4.1	261	21	AAC10644	Human secreted pro
c 44	33.8	4.1	1257	21	AAC74385	Human secreted pro
c 45	33.8	4.1	235033	19	AAV57926	Hereditary haemoch

ALIGNMENTS

RESULT 1
AAV62429
ID AAV62429 standard; cDNA; 820 BP.
XX
AC AAV62429;
XX
DT 30-DEC-1998 (first entry)
XX
DE Prostate cancer antigen (PCA3) cDNA splice variant 3.
XX
KW Prostate cancer antigen cDNA splice variant 3; PCA3; prostatic cancer;
KW PC; ds.
XX
OS Homo sapiens.
XX
PN WO9845420-A1.
XX
PD 15-OCT-1998.
XX
PF 09-APR-1998; 98WO-CA00346.
XX
PR 10-APR-1997; 97US-0041836.
XX
(DIAG-) DIAGNOCURE INC.
PA
XX
PI Bussemakers MJG;
XX
DR WPI; 1998-568347/48.
XX
PT New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
PT prevention and treatment of prostatic cancer
XX
PS Claim 4; Pages 77-78; 11lpp; English.
XX

CC The present sequence represents the prostate cancer antigen (PCA3)
CC cDNA splice variant 3 sequence comprising of exons 1, 3, and 4a
CC of the PCA3 gene. The PCA3 cDNA splice variant 3 sequence,
CC isolated from a human primary prostatic tumour tissue cDNA library,
CC was found in approximately 15% of the cDNA clones isolated. The
CC invention claims for PCA3 cDNA variants and the proteins they encode.
CC The invention also claims for antibodies against PCA3 protein. The
CC antibodies are claimed to be useful for detecting PCA3 protein in
CC immunosay tests, for diagnosing, assessing and prognosing of
CC prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin
CC or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed
CC to be useful for treating PC, while determining elevated levels of
CC PCA3 (as RNA or protein) is useful for detecting a predisposition
CC to development of PC, e.g. in prenatal tests. Detecting PCA3 protein
CC allows differentiation between malignant and benign prostatic disease,
CC and the level of PCA3 expression allows correlation with the grade of
CC tumour. PCA3 protein and its fragments are also claimed to be useful
CC in vaccines for preventing PC; in drug screens for identifying
CC specific (ant)agonists (potentially useful therapeutically) and for
CC studying protein-DNA interactions.
XX
SQ Sequence 820 BP; 262 A; 169 C; 191 G; 198 T; 0 other;

Query Match 100.0%; Score 820; DB 19; Length 820;
Best Local Similarity 100.0%; Pred. No. 3.3e-249;
Matches 820; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaagctggcatcagaaaaacagaggagatttggctcagccgaggagaccag 60
DB 1 agaagctggcatcagaaaaacagaggagatttggctcagccgaggagaccag 60

QY 61 gaagatctgcatgtgggaagaccctgatgatacagaggtgagaaataagaaggtgct 120
DB 61 gaagatctgcatgtgggaagaccctgatgatacagaggtgagaaataagaaggtgct 120

QY 121 gactttaccatctgagccacacatctgctgaaatggagataattaacatcactagaac 180
DB 121 gactttaccatctgagccacacatctgctgaaatggagataattaacatcactagaac 180

QY 181 agcaagatgacaataatgtctaaagttagtgacatgtttttgacatttccagcccttt 240
DB 181 agcaagatgacaataatgtctaaagttagtgacatgtttttgacatttccagcccttt 240

QY 241 aaatatccacacacaggaagcacaaaaaggaagcagagatccctgggagaaatgcc 300
DB 241 aaatatccacacacaggaagcacaaaaaggaagcagagatccctgggagaaatgcc 300

QY 301 ggcgcgcatttgggttcacgatgagcctgcctgtgctgtgctgtgctgtgagggaa 360
DB 301 ggcgcgcatttgggttcacgatgagcctgcctgtgctgtgctgtgctgtgagggaa 360

QY 361 ggcattagaaatgaatgatgtgttccttaagatggcagagatggcaggaacagatctgtt 420
DB 361 ggcattagaaatgaatgatgtgttccttaagatggcagagatggcaggaacagatctgtt 420

QY 421 gtgatatatttgaacgggtattcagatttgaatgaagtacacaaagtgcattacc 480
DB 421 gtgatatatttgaacgggtattcagatttgaatgaagtacacaaagtgcattacc 480

QY 481 aatgagagaaacacagacagaaaaatttgccttcacaaagacatgcacaaacaaa 540
DB 481 aatgagagaaacacagacagaaaaatttgccttcacaaagacatgcacaaacaaa 540

QY 541 tggaaactgtgatgacatgagcagccaaagctgggagagagataaccacggggcagag 600
DB 541 tggaaactgtgatgacatgagcagccaaagctgggagagagataaccacggggcagag 600

QY 601 gtcagattctgcccctgtcgttaactgtgcgttcataaaccacaaatcatttcatttc 660
DB 601 gtcagattctgcccctgtcgttaactgtgcgttcataaaccacaaatcatttcatttc 660

QY 661 taacctctaaacaaagctgtgttaatatctgatctctacggttctcttcttgggcccaaca 720

DB 661 taacctctaaacaaagctgtgttaatatctgatctctacggttctcttcttgggcccaaca 720

QY 721 ttctccatatatccagccacacacacatttttaatttagttccagatctgtactgtgac 780

DB 721 ttctccatatatccagccacacacacatttttaatttagttccagatctgtactgtgac 780

QY 781 cttctacactgtagataacatacattcattttgttcaaa 820

DB 781 cttctacactgtagataacatacattcattttgttcaaa 820

RESULT 2
AAV62428
ID AAV62428 standard; cDNA; 1872 BP.
XX
AC AAV62428;
XX
DT 30-DEC-1998 (first entry)
XX
DE Prostate cancer antigen (PCA3) cDNA splice variant 2.
XX
KW Prostate cancer antigen cDNA splice variant 2; PCA3; prostatic cancer;
KW PC; ds.
XX
OS Homo sapiens.
XX
PN WO9845420-A1.
XX
PD 15-OCT-1998.
XX
PF 09-APR-1998; 98WO-CA00346.
XX
PR 10-APR-1997; 97US-0041836.
XX
PA (DIAG-) DIAGNOCURE INC.
XX
PI Bussemakers MJG;
XX
DR WPI; 1998-568347/48.
XX
XX
XX New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
XX prevention and treatment of prostatic cancer
XX
XX Claim 4: Pages 76-77; 11pp; English.
XX
XX The present sequence represents the prostate cancer antigen (PCA3)
XX cDNA splice variant 2 sequence comprising of exons 1, 3, 4a and
XX 4b of the PCA3 gene. The PCA3 cDNA splice variant 2 sequence,
XX isolated from a human primary prostatic tumour tissue cDNA library,
XX was found in approximately 65% of the cDNA clones isolated. The
XX invention claims for PCA3 cDNA variants and the proteins they encode.
XX The invention also claims for antibodies against PCA3 protein. The
XX antibodies are claimed to be useful for detecting PCA3 protein in
XX immunosay tests, for diagnosing, assessing and prognosing of
XX prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin
XX or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed
XX to be useful for treating PC, while determining elevated levels of
XX PCA3 (as RNA or protein) is useful for detecting a predisposition
XX to development of PC, e.g. in prenatal tests. Detecting PCA3 protein
XX allows differentiation between malignant and benign prostatic disease,
XX and the level of PCA3 expression allows correlation with the grade of
XX tumour. PCA3 protein and its fragments are also claimed to be useful
XX in vaccines for preventing PC; in drug screens for identifying
XX specific (ant)agonists (potentially useful therapeutically) and for
XX studying protein-DNA interactions.
XX
SQ Sequence 1872 BP; 567 A; 389 C; 369 G; 539 T; 8 other;

Query Match 100.0%; Score 820; DB 19; Length 1872;
Best Local Similarity 100.0%; Pred. No. 5.3e-249;
Matches 820; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 aagaagctgcatcagaaaaacagaggagatttgtgtgctgcagccagggagaccag 60
DB 1 aagaagctgcatcagaaaaacagaggagatttgtgtgctgcagccagggagaccag 60
QY 61 gaagatctgcatcgttgggaagacctgatgatcacagaggtgagaaataagaaggctgt 120
DB 61 gaagatctgcatcgttgggaagacctgatgatcacagaggtgagaaataagaaggctgt 120
QY 121 gactttacatctgagggcacacatctgctgaaatggagataattaaacatcactagaaac 180
DB 121 gactttacatctgagggcacacatctgctgaaatggagataattaaacatcactagaaac 180
QY 181 agcaagatgacaataatgtctaaagttagtgacatgtttttgcacatttccagcccttt 240
DB 181 agcaagatgacaataatgtctaaagttagtgacatgtttttgcacatttccagcccttt 240
QY 241 aaatatccacacacacaggaagacaaaggagcacagagatccctgggagaaatgcc 300
DB 241 aaatatccacacacaggaagacaaaggagcacagagatccctgggagaaatgcc 300
QY 301 gcccaccatttgggtcatcgatgagcctgcctgtgcctgtccctgttccgcttggggaa 360
DB 301 gcccaccatttgggtcatcgatgagcctgcctgtgcctgttccgcttggggaa 360
QY 361 ggacattagaaaatgaattgatgttctctaaaggatgggcagagaaacagatcctgtt 420
DB 361 ggacattagaaaatgaattgatgttctctaaaggatgggcagagaaacagatcctgtt 420
QY 421 gtggatatttattgaacgggattacagatttgaatgaattgaattgaattgaattacc 480
DB 421 gtggatatttattgaacgggattacagatttgaatgaattgaattgaattgaattacc 480
QY 481 aatgagagaaaacagagagaaaattcttgatggtttcacaagacatgcacaacaaacaa 540
DB 481 aatgagagaaaacagagagaaaattcttgatggtttcacaagacatgcacaacaaacaa 540
QY 541 tggaaactgtgatgacatgagcgagcgaactgtgggagagataaacacggggcagagg 600
DB 541 tggaaactgtgatgacatgagcgagcgaactgtgggagagataaacacggggcagagg 600
QY 601 gtcaagattctggcctgcctgctaaactgtgcttcaataacaaatcatatttcatttc 660
DB 601 gtcaagattctggcctgcctgctaaactgtgcttcaataacaaatcatatttcatttc 660
QY 661 taacctcaaaaacaaagctgttgaatatctgacatctctacggttctctctgggccaaca 720
DB 661 taacctcaaaaacaaagctgttgaatatctgacatctctacggttctctctgggccaaca 720
QY 721 ttctccatatatccagccacactcatttttaatatatttagttccagatctgtactgtgac 780
DB 721 ttctccatatatccagccacactcatttttaatatatttagttccagatctgtactgtgac 780
QY 781 ctttctacactgtagaataaacattactcattttgttcaaa 820
DB 781 ctttctacactgtagaataaacattactcattttgttcaaa 820
```

RESULT 3

```
AAA06690/c
ID AAA06690 standard; cDNA; 812 BP.
XX
AC AAA06690;
XX
DT 13-JUN-2000 (first entry)
XX
DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:471.
XX
KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW immunogenic; cytostatic; vaccine; ss.
XX
OS Homo sapiens.
XX
```

```
PN WO200004149-A2.
XX
PD 27-JAN-2000.
XX
PF 14-JUL-1999; 99WO-US15838.
XX
PR 14-JUL-1998; 98US-0115453.
PR 14-JUL-1998; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX (CORI-) CORIXA CORP.
XX
PI Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;
XX WPI; 2000-171268/15.
DR
XX
PT New polypeptide useful for treating and diagnosing prostate cancer
XX comprises an immunogenic portion of prostate tumor protein -
PS Claim 1; Page 262; 263pp; English.
XX
CC The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate antibodies
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the polynucleotides encoding the polypeptides can be used as a probe or
CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC AAA82000 to AAA82020 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 812 BP; 198 A; 189 C; 169 G; 256 T; 0 other;

Query Match 99.0%; Score 812; DB 21; Length 812;
Best Local Similarity 100.0%; Pred. No. 1.1e-246;
Matches 812; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ctggcatcagaaaaacagaggaggatttgtgtgctgcagccgaggagacagaaaga 65
DB 812 CTGGCATCAGAAAAACAGAGGGGAGATTGTGTGGCTGCAGCCGAGGAGACGAGAGA 753
QY 66 tctgcatggtgggaaggacctgatgacagaggtgagaaataagaaggctgctgacct 125
DB 752 TCTGCATGTGGGAAGGACCTGATGATACAGAGGTGAGAAATAAGAAAGCGTCTGACTT 593
QY 126 taccatctgagggccacacatctgctgaaatggagataattaaacatcactagaaacagcaa 185
DB 692 TACCATCTGAGGCCACACATCTGCTGAATGGAGATAATTAACATCATTAGAACAGCAA 633
QY 186 gatgacaataataatgtctaaagttagtgacatgtttttgcacatttccagcccttaataa 245
DB 632 GATGACAATATATGTTCTTAAGTAGTACATGTGTGTGACATTTTCCAGCCCTTTAAATA 573
QY 246 tccacacacaggaagacacaaaggacacagagatccctgggagaaatcccgccgcg 305
DB 572 TCCACACACAGGAAGACAAAGGAAGACAGAGATCCCTGGGGAGAAATGCCCGGCCG 513
QY 306 ccattcttggtcatgatgagcctgcctgtgcctgttccctgttgcctgttggaggaagaca 365
DB 512 CCATCTTGGGTGTCATCGATGAGCCTCGCCCTGTGCCCTGTCCGCTGTGTGAGGAGGACA 453
QY 366 ttgaaaaatgaattgatgttctctaaaggatgggcagagaaacagatcctgtgtgga 425
DB 452 TTAGAAAATGAATTGATGTGTTCCTTAAAGGATGGGCAGGAAACAGATCCTCTTGTGGA 393
```

Qy	426	tatttatttgacgggattcacagatttgaaatgaagtcaacaagtgcattaccaatga	485
Db	392	TATTATTTTGAACGGGATTCAGATTTGAAATGAAGTCAAAAAGTGAGCATTTACCAATGA	333
Qy	486	gagaaaaacagacagagaaaaatcttgtggttcacaagacatgcaacaacaaaatggaa	545
Db	332	GAGAAAACAGACGAGAAAAATCTTGATGGCTTCACAAGACATGCACAAACAAAATGGAA	273
Qy	546	tactgtgatgcatagcaggcagaactggggagagagataaacacagggcgacagggccaq	605
Db	272	TACTGTGATGCATFAGGCGAGCCAAAGCTGGGGAGGAGATAACCACGGGCGACAGGGTCAG	213
Qy	606	gattctggccctgcctgctaactctgcgtctcataaccataacttatcttaacc	665
Db	212	GATTCTGGCCCTGCTGCCCTAAACTGTGGGTTCTAATACCAAATCATTTCTTAACC	153
Qy	666	ctcaaaaacaaagctgttgtaatatctgatctcacggtttccctctgggcccaactctc	725
Db	152	CTCAAAACAAAGCTGTGTGTAATATCTGATCTCTACGGTTCCTCTGGGGCCCAACATTCTC	93
Qy	726	cataataccagccacactcaatttttaatatatttagttcccagatctgtactgtgaaccttc	785
Db	92	CATATATCCAGCCACACTCATTTTTTAAATATTTAGTTAGTTCCCAGACTCTGTACTGTGACCTTTC	33
Qy	786	tacactgtagaataaacattactcatcttttggtc	817
Db	32	TACACTGTAGAATAAACATTACTCATTTTTTGGTC	1

RESULT	4
AAA06687	
ID	AAA06687 standard; cDNA; 3112 BP.
XX	
XX	AAA06687;
XX	
XX	
DT	13-JUN-2000 (first entry)
XX	
DE	Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:468.
XX	
XX	Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW	immunogenic; cytostatic; vaccine; ss.
KW	

XX	Homo sapiens.	
OS		
XX		
XX		
PN	WO200004149-A2.	
XX		
XX		
PD	27-JAN-2000.	
XX		
XX		
PF	14-JUL-1999;	99WO-US15838.
XX		
XX		
PR	14-JUL-1998;	98US-0115453.
PR	14-JUL-1998;	98US-0116134.
PR	23-SEP-1998;	98US-0159812.
PR	23-SEP-1998;	98US-0159822.
PR	15-JAN-1999;	98US-0232149.
PR	15-JAN-1999;	98US-0232880.
PR	09-APR-1999;	99US-0288946.
XX		
PA	(CORI-) CORIXA CORP.	
XX		
XX		
PI	Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;	
XX		
DR	WPI; 2000-171268/15.	
XX		
PT	New polypeptide useful for treating and diagnosing prostate cancer	
XX	comprises an immunogenic portion of prostate tumor protein -	
XX		
PS	Claim 1; Page 259-260; 263pp; English.	
XX		
CC	The present invention describes isolated polypeptides, comprising an	
CC	immunogenic portion of a prostate tumour protein (PTP). The polypeptides	
CC	and polynucleotides encoding them have cytostatic activity and can be	

used in vaccines and in gene therapy. The polypeptides and polynucleotides encoding them, antigen presenting cells which express the polypeptides, antibodies against the polypeptides and vaccines comprising them can be used for inhibiting the development of prostate cancer in a patient. The polypeptides can be used to generate antibodies or anti-idiotypic antibodies for passive immuno therapy. A portion of the polynucleotides encoding the polypeptides can be used as a probe or to modulate the expression of the polypeptides. AAA06241 to AAA06691 and AAA82000 to AAY92020 represent sequences used in the exemplification of the present invention.

Sequence 3112 BP: 975 A; 587 C; 624 G; 926 T; 0 other;

Query Match	88.5%	Score 725.8;	DB 21;	Length 3112;
Best Local Similarity	99.7%;	Pred. No. 4.5e-219;		
Matches 727;	Conservative	0; Mismatches 2;	Indels 0;	Gaps 0;
Qy	92	tacagaggtgagaataaagaagcgctgactttaccatctgagggcacacatctgctg	151	
Db	1307	taaatagggtgagaataaagaagcgctgactttaccatctgagggcacacatctgctg	1366	
Qy	152	aaatggagataattaaacatcatctagaacagcaagatgacaataataatgtctaaagtatg	211	
Db	1367	aaatggagataattaaacatcatctagaacagcaagatgacaataataatgtctaaagtatg	1426	
Qy	212	acatgtttttgcacatttccagccctttaaataatccacacacagaaacacaaaaagg	271	
Db	1427	acatgtttttgcacatttccagccctttaaataatccacacacagaaacacaaaaagg	1486	
Qy	272	aagcacagagatccctgggagaaatgccggccgcatcttgggtctcatctgatgagccctcg	331	
Db	1487	aagcacagagatccctgggagaaatgccggccgcatcttgggtctcatctgatgagccctcg	1546	
Qy	332	ccctgtgcctgggtcccgctgttgaggggaagacattagaaaaatgaattgatgtgttcctt	391	
Db	1547	ccctgtgcctgggtcccgctgttgaggggaagacattagaaaaatgaattgatgtgttcctt	1606	
Qy	392	aaagatgtggcagaaacacagatcctgttgggatattttttgaacgggattacagatt	451	
Db	1607	aaagatgtggcagaaacacagatcctgttgggatattttttgaacgggattacagatt	1666	
Qy	452	tgaatgaagtcaacaaagtgcagcattaccaatgagagaaaaacagacagaaaaatcttga	511	
Db	1667	tgaatgaagtcaacaaagtgcagcattaccaatgagagaaaaacagacagaaaaatcttga	1726	
Qy	512	tgggttcaacagacatgcacacacaaatggatactgtgatgcattgacatgagggcagccaag	571	
Db	1727	tgggttcaacagacatgcacacacaaatggatactgtgatgcattgacatgagggcagccaag	1786	
Qy	572	ctggggagaggagataaacacggggcagaggggtcaggatctctggccctgctgctctaaactgt	631	
Db	1787	ctggggagaggagataaacacggggcagaggggtcaggatctctggccctgctgctctaaactgt	1846	
Qy	632	gcgttcaaaccaaatcatttcataatttctaacctcacaacaaagctgttgtaatatct	691	
Db	1847	gcgttcaaaccaaatcatttcataatttctaacctcacaacaaagctgttgtaatatct	1906	
Qy	692	gatctctacggttctctctggggcccaaatcttccaatatccagccaacactcattttta	751	
Db	1907	gatctctacggttctctctggggcccaaatcttccaatatccagccaacactcattttta	1966	
Qy	752	atatttagttcccagatctgtactgtgaccttctacactgtagaataaacattactcatt	811	
Db	1967	atatttagttcccagatctgtactgtgaccttctacactgtagaataaacattactcatt	2026	
Qy	812	ttgttcaaa 820		
Db	2027	ttgttcaaa 2035		

RESULT 5
AAA06688/C

XX New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumor protein -
XX
XX
PS Claim 1; Page 261-262; 263pp; English.
XX
CC The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate antibodies
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the polynucleotides encoding the polypeptides can be used as a probe or
CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC AAY82000 to AAY82020 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 2426 BP; 717 A; 476 C; 548 G; 685 T; 0 other;

Query Match 88.5%; Score 725.6; DB 21; Length 2426;
Best Local Similarity 97.5%; Pred. No. 4.5e-219;
Matches 737; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 65 atctgcattggtggaagaccctgataacacagaggtgagaaataaagaagctgctgaact 124
|||||
Db 1802 ATTTGCTCTCAAAATGCTGATGATTTCCAGGTGAGAAATGAAGAAAGGCTGCTGACT 1743
|||||
QY 125 ttaccatctgagggccacacatctgctgaaatggagataattaacatcactagaaacagca 184
|||||
Db 1742 TTACCATCTGAGGCCACACATCTGCTGAATGGAGATAATTAACATCCTAGAAACAGCA 1683
|||||
QY 185 agatgacaataatagtcttaagttagtgacatgtttttgacacatttccagccctttaaatt 244
|||||
Db 1682 AGATGACAATATATGCTTAAGTAGTGACATGTTTTTGGACATTTCCAGCCCTTTAAAT 1623
|||||
QY 245 atccacacacaggaagacacaaaaggaacacagagatccctggagaaaatcccggcc 304
|||||
Db 1622 ATCCACACACAGGAAGACAAAAGAACACAGAGATCCCTGGGAGAAATGCCGGCC 1563
|||||
QY 305 gccatctgggtcatgatgagcctgcctgtgcctgtgctcctgttgagggaaaggac 364
|||||
Db 1562 GCCATCTGGGTCATGATGAGCTCGCCCTGTGCTGTGCTGCTGCTGCTGCTGCTGCTG 1503
|||||
QY 365 attagaaaatgaattgatgtgttccttaagagatggcgaggaacacagatcctgtgtgtg 424
|||||
Db 1502 ATTAGAAAATGAATTGATGTGTTCTTAAAGGATGGGCGAGAAAACAGATCCTGTTGTG 1443
|||||
QY 425 atatttattgaaacgggattacagattgaaatgaagtacaaagtgcattaccattg 484
|||||
Db 1442 ATATTATTATTTACGGGATTACAGATTTGAAATGAAGTCAAAAGTGAGCATTTACCAATG 1383
|||||
QY 485 agaggaacacagacgagaaatcttgatggcttcacaaagacatgcacaaacaaaatgga 544
|||||
Db 1382 AGAGAAAACAGACGAGAAAATCTTGATGCTTTCACAAAGACATGCAACAAACAAATGGA 1323
|||||
QY 545 atactgtgatgacatgagggcagcaagctgggagagataaacacaggggcagagggtca 604
|||||
Db 1322 ATACTGTGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1263
|||||
QY 605 ggattctggcctgctgcctaaactgctgcttataaccacaaatcatttcatttctaac 664
|||||
Db 1262 GGATTCTGGCCCTGCTGCTTAACCTGCTTCTATACCAAAATCATTTATATTCTAAC 1203
|||||
QY 665 cctcaaaacaaagctgttgtaatatctgatctctacaggttcctctgggcccacattct 724
|||||
Db 1202 CCTCAAAACAAAGCTGTTGTAATCTGATCTCTACGGTTCCTTCTGGGCCCAACATCT 1143
|||||
QY 725 coatatatccagccacactcatttttaataatttagttccagatctgtactgtgaacttt 784
|||||

Db 1142 CCATATATCCAGCCACACTCAATTTTAATATTTAGTTCACAGATCTGTACTGTGACCTTT 1083
QY 785 ctacactgtagaataacattactcatttctgttcaaa 820
|||||
Db 1082 CTACACTGTAGAATAACATTACTCAITTTGTTTCAAA 1047
|||||

RESULT 7
AAV62427
ID AAV62427 standard; cDNA; 2037 BP.
XX
AC AAV62427;
XX
DT 30-DEC-1998 (first entry)
XX
DE Prostate cancer antigen (PCA3) cDNA splice variant 1.
XX
KW Prostate cancer antigen cDNA splice variant 1; PCA3; prostatic cancer;
KW PC; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 379..534
FT /tag= a
FT /product= "PCA3 protein variant 1"
FT polyA_signal 2019..2024
FT /*tag= b
XX PN WO9845420-A1.
XX
PD 15-OCT-1998.
XX
PF 09-APR-1998; 98WO-CA00346.
XX
PR 10-APR-1997; 97US-0041836.
XX
PA (DIAG-) DIAGNOCURE INC.
XX
PI Bussemakers MJG;
XX
DR WPI; 1998-568347/48.
DR P-PSDB; AAW79736.
XX
XX New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
PT prevention and treatment of prostatic cancer
XX
PS Claim 3; Fig 2B-2J; lllpp; English.
XX
CC The present sequence represents the prostate cancer antigen (PCA3)
CC cDNA splice variant 1 sequence comprising of exons 1, 2, 3, 4a and
CC 4b of the PCA3 gene. The PCA3 cDNA splice variant 1 sequence,
CC isolated from a human primary prostatic tumour tissue cDNA library,
CC was found in approximately 5% of the cDNA clones isolated. The
CC invention claims for PCA3 cDNA variants and the proteins they encode.
CC The invention also claims for antibodies against PCA3 protein. The
CC antibodies are claimed to be useful for detecting PCA3 protein in
CC immunoassay tests, for diagnosing, assessing and prognosing of
CC prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin
CC or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed
CC to be useful for treating PC, while determining elevated levels of
CC PCA3 (as RNA or protein) is useful for detecting a predisposition
CC to development of PC, e.g. in prenatal tests. Detecting PCA3 protein
CC allows differentiation between malignant and benign prostatic disease,
CC and the level of PCA3 expression allows correlation with the grade of
CC tumour. PCA3 protein and its fragments are also claimed to be useful
CC in vaccines for preventing PC; in drug screens for identifying
CC specific (ant)agonists (potentially useful therapeutically) and for
CC studying protein-DNA interactions.
XX
SQ Sequence 2037 BP; 622 A; 426 C; 406 G; 575 T; 8 other;

Query Match 88.5%; Score 725.4; DB 19; Length 2037;
Best Local Similarity 99.9%; Pred. No. 4.7e-219;
Matches 726; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	94	cagaggtgagaaataaagaaggctgctgactttaccatctgagggccacacatctgctgaa	153
Db	259	caggggtgagaaataaagaaggctgctgactttaccatctgagggccacacatctgctgaa	318
Qy	154	atggagataataaacatcactagaaacagcaagatgacaataataatgtctgaagtagtac	213
Db	319	atggagataataaacatcactagaaacagcaagatgacaataataatgtctgaagtagtac	378
Qy	214	atgttttgcacatttcagcccttttaatatccacacacaggaagacacaaaggaa	273
Db	379	atgttttgcacatttcagcccttttaatatccacacacaggaagacacaaaggaa	438
Qy	274	gcacagagatccctgggagaaatcccggcgccatctctgggtcatcgatgagcctgcc	333
Db	439	gcacagagatccctgggagaaatcccggcgccatctctgggtcatcgatgagcctgcc	498
Qy	334	ctgtgcttgcctccgctgtgaggaagacattagaaaaatgaattgattgttcttaa	393
Db	499	ctgtgcttgcctccgctgtgaggaagacattagaaaaatgaattgattgttcttaa	558
Qy	394	aggatggcgagaaacagatcctgttggatattttgaacggattacagatttg	453
Db	559	aggatggcgagaaacagatcctgttggatattttgaacggattacagatttg	618
Qy	454	aatgaagtcaaaagttagcattaccatgagaggaagaaacagacagaaatcttgatg	513
Db	619	aatgaagtcaaaagttagcattaccatgagaggaagaaacagacagaaatcttgatg	678
Qy	514	gcttcacagacatgcacaaacaaaatggaaatctgtgacatgagcagccaagct	573
Db	679	gcttcacagacatgcacaaacaaaatggaaatctgtgacatgagcagccaagct	738
Qy	574	ggggaggagataaacacaggggcagaggtcaggattctggccctgctgcctaaactgtgc	633
Db	739	ggggaggagataaacacaggggcagaggtcaggattctggccctgctgcctaaactgtgc	798
Qy	634	gttcataaccataatatttcatttccatccatccatccatccatccatccatccatccat	693
Db	799	gttcataaccataatatttcatttccatccatccatccatccatccatccatccatccat	858
Qy	694	tctctacggttctcttggggccacatttccatccatccatccatccatccatccatccat	753
Db	859	tctctacggttctcttggggccacatttccatccatccatccatccatccatccatccat	918
Qy	754	atttagttccagatctgactgtgacatttctacactgtagaataaacattactattt	813
Db	919	atttagttccagatctgactgtgacatttctacactgtagaataaacattactattt	978
Qy	814	gttcaaa 820	
Db	979	gttcaaa 985	

RESULT 8
AAV62430
ID AAV62430 standard; cDNA; 3582 BP.
XX
AC AAV62430;
XX
DT 30-DEC-1998 (first entry)
XX
DE Prostate cancer antigen (PCA3) wild-type cDNA.
XX
KW Prostate cancer antigen cDNA; PCA3; prostatic cancer;
KW PC; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers

CDS 401..556
FT /*tag= a
FT polyA_signal 983..987
FT /*tag= b
FT polyA_signal 2041..2046
FT /*tag= c
FT polyA_signal 2597..2602
FT /*tag= d
FT polyA_signal 3494..3496
FT /*tag= e
XX
PN WO9845420-A1.
XX
PD 15-OCT-1998.
XX
PF 09-APR-1998; 98WO-CA00346.
XX
PR 10-APR-1997; 97US-0041836.
XX
(DIAG-) DIAGNOCURE INC.
XX
PI Bussemakers MJG;
XX
DR WPI; 1998-568347/48.
DR P-PSDB; AAW79738.
XX
PT New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
prevention and treatment of prostatic cancer
XX
PS Claim 3; Fig 5B-5F; 11lpp; English.
XX
CC The present sequence represents the prostate cancer antigen (PCA3)
wild-type cDNA sequence comprising of exons 1, 2, 3, 4a-4d of the
PCA3 gene. The invention claims for PCA3 cDNA variants and the
proteins they encode. The invention also claims for antibodies
against PCA3 protein. The antibodies are claimed to be useful for
detecting PCA3 protein in immunoassay tests, for diagnosing, assessing
and prognosing of prostatic cancer (PC). Antibodies, optionally
coupled to a cytotoxin or radioisotope, and nucleic acids antisense
to PCA3 cDNA are claimed to be useful for treating PC, while determining
elevated levels of PCA3 (as RNA or protein) is useful for detecting a
predisposition to development of PC, e.g. in prenatal tests. Detecting
PCA3 protein allows differentiation between malignant and benign
prostatic disease, and the level of PCA3 expression allows correlation
with the grade of tumour. PCA3 protein and its fragments are also
claimed to be useful in vaccines for preventing PC; in drug screens
for identifying specific (ant)agonists (potentially useful
therapeutically) and for studying protein-DNA interactions.
XX
SQ Sequence 3582 BP; 1052 A; 788 C; 679 G; 1063 T; 0 other;

Query Match 88.5%; Score 725.4; DB 19; Length 3582;
Best Local Similarity 99.9%; Pred. No. 6.5e-219;
Matches 726; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 94 cagaggtgagaaataaagaaggctgctgactttaccatctgagggccacacatctgctgaa 153
Db 281 caggggtgagaaataaagaaggctgctgactttaccatctgagggccacacatctgctgaa 340
Qy 154 atggagataataaacatcactagaaacagcaagatgacaataataatgtctgaagtagtac 213
Db 341 atggagataataaacatcactagaaacagcaagatgacaataataatgtctgaagtagtac 400
Qy 214 atgttttgcacatttcagcccttttaatatccacacacaggaagacacaaaggaa 273
Db 401 atgttttgcacatttcagcccttttaatatccacacacaggaagacacaaaggaa 460
Qy 274 gcacagagatccctgggagaaatcccggcgccatctctgggtcatcgatgagcctgcc 333
Db 461 gcacagagatccctgggagaaatcccggcgccatctctgggtcatcgatgagcctgcc 520

```
QY 334 ctgtgctgtcccgctgtgtgagggaagacattagaaaaatgaattgattgttctcttaa 393
Db 521 ctgtgctgtcccgctgtgtgagggaagacattagaaaaatgaattgattgttctcttaa 580
QY 394 aggatggcaggaaacagatcctgtgttgatatttatttgaacgggattacagatttg 453
Db 581 aggatggcaggaaacagatcctgtgttgatatttatttgaacgggattacagatttg 640
QY 454 aaatgaagtccaaagtgaattaccattacaaatgagagaaacagacagaaaaatcttgatg 513
Db 641 aaatgaagtccaaagtgaattaccattacaaatgagagaaacagacagaaaaatcttgatg 700
QY 514 gttccaaagacatgcaacaaacaaatggaatactgtgatgacatgagcgagcgaagct 573
Db 701 gttccaaagacatgcaacaaacaaatggaatactgtgatgacatgagcgagcgaagct 760
QY 574 ggggaggagataaacacggggcagaggggtcagattctgagctgtgctgctaaactgtgc 633
Db 761 ggggaggagataaacacggggcagaggggtcagattctgagctgtgctgctaaactgtgc 820
QY 634 gttcaataccaaatcatttcatttcttaacccctcaaaacaaagctgttgttaattctga 693
Db 821 gttcaataccaaatcatttcatttcttaacccctcaaaacaaagctgttgttaattctga 880
QY 694 tctctacggttcttctgtggcccaacattctccatataatccagccacactattttta 753
Db 881 tctctacggttcttctgtggcccaacattctccatataatccagccacactattttta 940
QY 754 attagttcccagatctgactgtacattcttctacactgtacactgtagaataactactatt 813
Db 941 attagttcccagatctgactgtacattcttctacactgtagaataactactatt 1000
QY 814 gttcaaa 820
Db 1001 gttcaaa 1007

RESULT 9
AA37486
ID AAX37486 standard; cDNA; 597 BP.
XX
AC AAX37486;
XX
DT 06-JUL-1999 (first entry)
DE Human secreted protein cDNA fragment containing gene 36.
XX
KW Human; secreted protein; treatment; prevention; protein therapy; AIDS;
KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW immune system disease; autoimmune disease; hepatic disease; lymphoma;
KW renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
KW cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
KW pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
KW arthritis; malignancy; digestive; endocrine; infection; ss.
XX
OS Homo sapiens.
XX
PN WO918208-A1.
XX
PD 15-APR-1999.
XX
PF 01-OCT-1998; 98WO-US20775.
XX
PR 02-OCT-1997; 97US-0060884.
PR 02-OCT-1997; 97US-0060833.
PR 02-OCT-1997; 97US-0060836.
PR 02-OCT-1997; 97US-0060837.
PR 02-OCT-1997; 97US-0060838.
PR 02-OCT-1997; 97US-0060839.
PR 02-OCT-1997; 97US-0060843.
PR 02-OCT-1997; 97US-0060862.
PR 02-OCT-1997; 97US-0060866.
```

```
PR 02-OCT-1997; 97US-0060874.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Carter KC, Duan DR, Endress GA, Feng P, Ferrie AM;
PI Florence KA, Greene JM, Janat F, Lafleur DW, Ni J;
PI Rosen CA, Ruben SM, Shi Y, Young P, Yu G;
XX
WPI: 1999-264022/22.
DR P-PSDB; AAY07887.
XX
PT New isolated human genes and the secreted polypeptides they encode
PS Claim 1a; Page 247; 368pp; English.
XX
CC This invention describes novel isolated human genes and the secreted
CC proteins they encode. The products of the invention are useful for
CC preventing, treating or ameliorating medical conditions, e.g. by protein
CC or gene therapy. Also pathological conditions can be diagnosed by
CC determining the amount of the new polypeptides in a sample or by
CC determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 101 polynucleotides, based
CC on which tissues they are most highly expressed in, and include
CC developing products for the diagnosis or treatment of cancer, tumours,
CC neurodegenerative disorders, developmental abnormalities and fetal
CC deficiencies, blood disorders, leukemias, diseases of the immune system,
CC autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,
CC allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate
CC disease, skeletal or cardiac muscle disorders, pulmonary disorders,
CC transplant rejection, disorders involving osteoclasts such as
CC osteoporosis, arthritis or malignancies, digestive/endocrine disorders,
CC infections and AIDS. The human secreted proteins of the invention are
CC represented in AAY07852-Y07993 and the encoding nucleic acids are
CC represented in AAX37451-X37552.
XX
SQ Sequence 597 BP; 181 A; 131 C; 134 G; 150 T; 1 other;

Query Match 63.9%; Score 523.6; DB 20; Length 597;
Best Local Similarity 99.8%; Pred. No. 1.6e-155;
Matches 523; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 297 gccggcgccatcttggttcacatgagcctgcctgcctgtgctggtcccgctgtgag 356
Db 54 gccggcgccatcttggttcacatgagcctgcctgcctgtgctggtcccgctgtgag 113
QY 357 ggaaggaattagaaatgaattgttcttaaaagatggcgagaaacagatcc 416
Db 114 ggaaggacattagaaatgaattgttcttaaaagatggcgagaaacagatcc 173
QY 417 tttgtggatattttgaacgggattacagatttgaatgaattgaatgaatgagcat 476
Db 174 tttgtggatattttgaacgggwttaacagatttgaatgaatgaatgaatgagcat 233
QY 477 taccatgagagaaacagacagaaaaatcttgatggcttcacagacatgcaacaac 536
Db 234 taccatgagagaaacagacagaaaaatcttgatggcttcacagacatgcaacaac 293
QY 537 aaaaatggaatactgtgatgacatgagcgaacgaatggggaggagataaacacgggca 596
Db 294 aaaaatggaatactgtgatgacatgagcgaacgaatggggaggagataaacacgggca 353
QY 597 gaggttcaggattctggccctgctgctaaactgtgcttcaataacaaatcattcata 656
Db 354 gaggttcaggattctggccctgctgctaaactgtgcttcaataacaaatcattcata 413
QY 657 ttctaacctcaaaacaaagctgttgaatctgtatctacgttctctcttctggccc 716
Db 414 ttctaacctcaaaacaaagctgttgaatctgtatctacgttctctcttctggccc 473
QY 717 aacattctccatataatccagccacactatttttaattagttccagactgtactg 776
Db 474 aacattctccatataatccagccacactatttttaattagttccagactgtactg 533
```

```
QY 777 tgacctttctacactgtagaataacattactactcttttggttcaaa 820
|||||
Db 534 tgacctttctacactgtagaataacattactactcttttggttcaaa 577
|||||

RESULT 10
AAA06545
ID AAA06545 standard; cDNA; 718 BP.
XX
AC AAA06545;
XX
DT 13-JUN-2000 (first entry)
XX
DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:313.
XX
KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW immunogenic; cytostatic; vaccine; ss.
XX
OS Homo sapiens.
PN WO200004149-A2.
XX
PD 27-JAN-2000.
PF 14-JUL-1999; 99WO-US15838.
PR 14-JUL-1998; 98US-0115453.
PR 14-JUL-1998; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX
PA (CORI-) CORIXA CORP.
XX
PI Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;
XX WPI; 2000-171268/15.
XX
PT New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumor protein -
XX
PS Claim 1; Page 199-200; 263pp; English.
XX
CC The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate antibodies
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the polynucleotides encoding the polypeptides can be used as a probe or
CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC AA82000 to AA82020 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 718 BP; 222 A; 145 C; 169 G; 172 T; 10 other;

Query Match 62.6%; Score 513.2; DB 21; Length 718;
Best Local Similarity 97.3%; Pred. NO. 3.4e-152;
Matches 585; Conservative 0; Mismatches 9; Indels 7; Gaps 6;

QY 27 ggagattgtgtg-cgcagccgaggagaccaggaagatctgcatgtgtggaaggacc 85
|||||
Db 1 ggagattgtgtgttcagcccgaggagaccaggaagatctgcatgtgtggaaggacc 60
|||||
QY 86 tgatgatacagaggtgagaataaagaaggctgctgactttaccatctgaggccacacat 145
|||||
```

XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.
 XX
 SQ Sequence 437 BP; 140 A; 87 C; 118 G; 88 T; 4 other;

Query Match 50.5%; Score 414.2; DB 21; Length 437;
 Best Local Similarity 99.0%; Pred. No. 5.2e-121;
 Matches 413; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 agaagctgcatcagaaaaacagaggggagattgtgtgctgcagcccgaggagaccag 60
 |||||
 Db 21 agaagctgcatcagaaaaacagaggggagattgtgtgctgcagcccgaggagaccag 80
 |||||

QY 61 gaagatctcgtgtgtggaagaccctgatgacagagtgagaaataagaagcctgct 120
 |||||
 Db 81 gaagatctcgtgtgtggaagaccctgatgacagagtgagaaataagaagcctgct 140
 |||||

QY 121 gactttaccatctgagggccacacatctgctgaaatggagataattaacatcactagaaac 180
 |||||
 Db 141 gactttaccatctgagggccacacatctgctgaaatggagataattaacatcactagaaac 200
 |||||

QY 181 agcaagatgacaataataatgtctaagtagtagacatgtttttcacatttccagcccttt 240
 |||||
 Db 201 agcaagatgacaataataatgtctaagtagtagacatgtttttcacatttccagcccttt 260
 |||||

QY 241 aaatatccacacacacaggaagacacaaaaggagacacagagatccctgggagaaatgcc 300
 |||||
 Db 261 aaatatccacacacacaggaagacacaaaaggagacacagagatccctgggagaaatgcc 320
 |||||

QY 301 ggcggccattctgggtatcgatgagcctcgccctgtgctgctgctgctgctgctgctgct 360
 |||||
 Db 321 ggcggccattctgggtatcgatgagcctcgccctgtgctgctgctgctgctgctgctgct 380
 |||||

QY 361 gacattagaaatgaattgattgttcttaaggatggcgaggagaaacagatcct 417
 |||||
 Db 381 ggcattagaaatgrattgtgtgtcttaaggatggcgaggagaaacagatcct 437
 |||||

RESULT 12

AAZ33445
 ID AAZ33445 standard; CDNA; 359 BP.

XX
 AC AAZ33445;

XX
 DT 08-DEC-1999 (first entry)

XX
 DE Human prostate cancer-associated EST 23.

XX
 KW Expressed sequence tag; EST; prostate tumor; antitumor; treatment;

XX
 KW gene therapy; tissue specificity human; ss.

XX
 OS Homo sapiens.

XX
 PN DE19811193-A1.

XX
 PD 16-SEP-1999.

XX
 PF 10-MAR-1998; 98DE-1011193.

XX
 KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;

PR 10-MAR-1998; 98DE-1011193.
 XX
 PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
 PI Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dahl E, Rosenthal A;
 XX
 DR WPI; 1999-519628/44.
 DR P-PSDB; AAY48243.
 XX
 PT New nucleic acid expressed at high level in prostatic tumor tissue and
 PT encoded polypeptides, useful for treating cancer and screening for
 PT therapeutic agents -
 XX
 PS Claim 1a; 87; 166pp; German.
 XX
 CC This invention describes novel nucleic acid sequences (A) that are
 CC expressed at high level in prostatic tumor tissue and encode gene
 CC products or their fragments. The products of the invention have
 CC antitumor activity. Polypeptides (I) encoded by (A) are used: (i) for
 CC identifying agents for treatment of prostatic cancer and (ii) for
 CC therapy of prostate cancer, optionally where expressed by gene therapy
 CC methods. (A) is also used to isolate full-length genes (for gene therapy)
 CC and for recombinant production of (I), which can be used to raise
 CC specific antibodies. (A) are identified by assembly of ESTs (expressed
 CC sequence tags) before they are analyzed for expression pattern (tissue
 CC specificity). This approach eliminates many of the false results, as
 CC regards tissue specificity, associated with known methods that use
 CC single (usually short) ESTs. AAZ33423-233476 represent expressed
 CC sequence tags described in the method of the invention.
 XX
 SQ Sequence 359 BP; 121 A; 75 C; 94 G; 69 T; 0 other;

Query Match 40.2%; Score 330; DB 20; Length 359;
 Best Local Similarity 100.0%; Pred. No. 2e-94;
 Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaagctgcatcagaaaaacagaggggagattgtgtgctgcagcccgaggagaccag 60
 |||||
 Db 30 agaagctgcatcagaaaaacagaggggagattgtgtgctgcagcccgaggagaccag 89
 |||||

QY 61 gaagatctcgtgtgtggaagaccctgatgacagagtgagaaataagaagcctgct 120
 |||||
 Db 90 gaagatctcgtgtgtggaagaccctgatgacagagtgagaaataagaagcctgct 149
 |||||

QY 121 gactttaccatctgagggccacacatctgctgaaatggagataattaacatcactagaaac 180
 |||||
 Db 150 gactttaccatctgagggccacacatctgctgaaatggagataattaacatcactagaaac 209
 |||||

QY 181 agcaagatgacaataataatgtctaagtagtagacatgtttttcacatttccagcccttt 240
 |||||
 Db 210 agcaagatgacaataataatgtctaagtagtagacatgtttttcacatttccagcccttt 269
 |||||

QY 241 aaatatccacacacacaggaagacacaaaaggagacacagagatccctgggagaaatgcc 300
 |||||
 Db 270 aaatatccacacacacaggaagacacaaaaggagacacagagatccctgggagaaatgcc 329
 |||||

QY 301 ggcggccattctgggtatcgatgagcctc 330
 |||||
 Db 330 ggcggccattctgggtatcgatgagcctc 359
 |||||

RESULT 13

AAA06520/C

ID AAA06520 standard; CDNA; 301 BP.

XX
 AC AAA06520;

XX
 DT 13-JUN-2000 (first entry)

XX
 DE Human immunogenic prostate tumour protein CDNA sequence SEQ ID NO:287.

XX
 KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;

KW immunogenic; cytostatic; vaccine; ss.

XX Homo sapiens.

PN WO200004149-A2.

XX 27-JAN-2000.

XX 14-JUL-1999; 99WO-US15838.

XX 14-JUL-1998; 98US-0115453.

PR 14-JUL-1998; 98US-0116134.

PR 23-SEP-1998; 98US-01159812.

PR 23-SEP-1998; 98US-01159822.

PR 15-JAN-1999; 99US-0232149.

PR 15-JAN-1999; 99US-0232880.

PR 09-APR-1999; 99US-0288946.

XX (CORI-) CORIXA CORP.

FA Dillion DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;
PI WPI; 2000-171268/15.

XX New polypeptide useful for treating and diagnosing prostate cancer

XX comprises an immunogenic portion of prostate tumor protein -
PT Claim 1; Page 192; 263pp; English.

XX The present invention describes isolated polypeptides, comprising an

CC immunogenic portion of a prostate tumor protein (ptp). The polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate antibodies
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the polynucleotides encoding the polypeptides can be used as a probe or
CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC AA82000 to AA82020 represent sequences used in the exemplification of
CC the present invention.

XX Sequence 301 BP; 76 A; 58 C; 70 G; 97 T; 0 other;

Query Match 35.28; Score 288.4; DB 21; Length 301;

Best Local Similarity 99.38; Pred. No. 2.6e-81;

Matches 300; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 472 agcattaccatgagagaaacagacagagaaatcttgatggtctcacaagacatgcaa 531

Db 301 AGCATTACCATGAGAGAAACACAGACAGAAATCTTGATGGCTTCACAGACATGCAA 242

Qy 532 caaacaataatggaatactgtgatgacatgagcagcgaagctgggagagataaccacg 591

Db 241 CAACAAATGGAATGTAATGTAACATGAGCGAGCCAAAGCTGGGAGGAGATAACACG 182

Qy 592 ggccagaggggcagattctggcctgctgctaaactgtcggttcataaccacaatcatt 651

Db 181 GGGCAGAGGGTCAGAGTTCTGGCCCTGCTGCTAAACGTGCTTCAATACCAATCAT 122

Qy 652 tcatattctaacctcacaagaagctgtgttaatactgattctctacgttctctctg 711

Db 121 TCATATTTCTAACCTCAAAACAAAGCTGTGTATATATCTGATCTCTAC-GTTCCTCTG 63

Qy 712 ggcacaactcttcataatccagccacactcatttttaattagttccagatctg 771

Db 62 GGCACCAACTTCTCATATATCCAGCCACACTCATTTTTTATATATTAGTTCCTCCAGATCTG 3

Qy 772 ta 773

Db 2 TA 1

RESULT 14

AAV33791

ID AAV33791 standard; cDNA; 123 BP.

XX AAV33791;

XX 30-DEC-1998 (first entry)

XX Prostate cancer antigen (PCA3) gene transcription start site.

XX Prostate cancer antigen gene transcription start site; PCA3;

XX prostatic cancer; PC; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT misc_feature 122..123

FT /*tag= a

FT /note= "Splice acceptor site"

XX WO9845420-A1.

XX 15-OCT-1998.

XX 09-APR-1998; 98WO-CA00346.

XX 10-APR-1997; 97US-0041836.

XX (DIAG-) DIAGNOCURE INC.

XX Bussemakers MJG;

XX WPI; 1998-568347/48.

XX New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
XX prevention and treatment of prostatic cancer
XX Example 3; Fig 4; 11pp; English.
XX The present sequence represents a transcription start site sequence
XX of the prostate cancer antigen (PCA3) gene. The invention claims for
XX PCA3 cDNA variants and the proteins they encode. The invention also
XX claims for antibodies against PCA3 protein. The antibodies are
XX claimed to be useful for detecting PCA3 protein in immunoassay tests,
XX for diagnosing, assessing and prognosing of prostatic cancer (PC).
XX Antibodies, optionally coupled to a cytotoxin or radioisotope, and
XX nucleic acids antisense to PCA3 cDNA are claimed to be useful for
XX treating PC, while determining elevated levels of PCA3 (as RNA or
XX protein) is useful for detecting a predisposition to development of
XX PC, e.g. in prenatal tests. Detecting PCA3 protein allows
XX differentiation between malignant and benign prostatic disease,
XX and the level of PCA3 expression allows correlation with the grade of
XX tumour. PCA3 protein and its fragments are also claimed to be useful
XX in vaccines for preventing PC; in drug screens for identifying
XX specific (ant)agonists (potentially useful therapeutically) and for
XX studying protein-DNA interactions.

XX Sequence 123 BP; 41 A; 19 C; 45 G; 18 T; 0 other;

Query Match 12.28; Score 100; DB 19; Length 123;
Best Local Similarity 100.08; Pred. No. 6.1e-22;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 agaagctggcatcagaaaaacagaggagagatttggctgcagccgagggagaccag 60

Db 24 agaagctggcatcagaaaaacagaggagagatttggctgcagccgagggagaccag 83

Qy 61 gaagatctgcatggtgggaagaccctgatgatacagagt 100

Db 84 gaagatctgcatggtgggaagaccctgatgatacagagt 123

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 08:05:28 ; Search time 10902.7 Seconds
(without alignments)
1163.342 Million cell updates/sec

Title: US-09-402-713A-4
Perfect score: 820
Sequence: 1 agaagctgcatcagaaaa.....cattactcattgttcaaa 820

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: gb_ba1:*
17: gb_ba2:*
18: gb_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vl:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vl1:*
59: gb_vl2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	725.4	88.5	3923	88	AF103907 Homo sapi
2	542.4	66.1	5435	88	AF103908 Homo sapi
3	542.4	66.1	164371	80	AL390239 Homo sapi
	542.4	66.1	173831	80	AL359314 Homo sapi
c 5	542.4	66.1	267581	80	AL358573 Homo sapi
6	330	40.2	359	9	AX018075 Sequence
7	100.2	12.2	143675	79	AL161625 Homo sapi
8	100.2	12.2	267581	80	AL358573 Homo sapi

```
c 9 98.6 12.0 172298 61 AC009556
c 10 98 12.0 580 89 AF279290
c 11 57.4 7.0 7218 10 I66494
c 12 45 5.5 191319 63 AC015653
c 13 42.8 5.2 9131 85 AB038781
c 14 42.6 5.2 8912 85 AB038782
c 15 42.6 5.2 10708 89 AF113616
c 16 42 5.1 175144 73 AC068502
c 17 42 5.1 199103 76 AC080076
c 18 41.8 5.1 2458 94 MMSTRALGN
c 19 40.8 5.0 115583 93 HSJ351K20
c 20 40.8 5.0 164478 83 AF002450
c 21 39.8 4.9 141492 73 AC068941
c 22 39.8 4.9 175353 73 AC069061
c 23 39.8 4.9 176958 74 AC069123
c 24 39.6 4.8 76175 15 TIN6
c 25 39.4 4.8 173547 79 AL162716
c 26 39.2 4.8 50883 90 AL359708
c 27 39 4.8 993 94 MMSNRPN3
c 28 39 4.8 123004 92 HS196823
c 29 39 4.8 175355 64 AC016480
c 30 39 4.8 183259 76 AC079914
c 31 38.8 4.7 40998 86 AC005503
c 32 38.8 4.7 41730 93 HSAC000372
c 33 38.8 4.7 129290 66 AC020761
c 34 38.8 4.7 145050 70 AC026679
c 35 38.8 4.7 178256 70 AC026878
c 36 38.8 4.7 190778 92 HS142L7
c 37 38.6 4.7 193930 86 AC006296
c 38 38.6 4.7 196807 70 AC028602
c 39 38.6 4.7 210348 69 AC025845
c 40 38.4 4.7 8589 93 HSLPLINT
c 41 38.4 4.7 17427 93 HSA6976
c 42 38.4 4.7 42013 97 HSV698D2
c 43 38.4 4.7 142084 84 HS540A13
c 44 38.4 4.7 147055 92 HS372K1
c 45 38.4 4.7 171430 70 AC027160
```

ALIGNMENTS

```
RESULT 1
AF103907
LOCUS AF103907 3923 bp mRNA PRI 14-AUG-2000
DEFINITION Homo sapiens non-coding RNA DD3 sequence.
ACCESSION AF103907
VERSION AF103907.1 GI:6165973
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3923)
AUTHORS Bussemakers,M.J.G., van Bokhoven,A., Verhaegh,G.W., Smit,F.P.,
Karthaas,H.F., Schalken,J.A., Debruyne,F.M., Ru.N. and Isaacs,W.B.
TITLE DD3: a new prostate-specific gene, highly overexpressed in prostate
cancer
JOURNAL Cancer Res. 59 (23), 5975-5979 (1999)
MEDLINE 20072260
PUBMED 10606244
REFERENCE 2 (bases 1 to 3923)
AUTHORS Bussemakers,M.J.G., Van Bokhoven,A., Verhaegh,G.W., Smit,F.P.,
Karthaas,H.F.M., Schalken,J.A., Debruyne,F.M.J., Ru.N. and
Isaacs,W.B.
TITLE Direct Submission
JOURNAL Submitted (28-OCT-1998) Urology Research Laboratory, University
Hospital Nijmegen, P.O. Box 9101, Nijmegen 6500 HB, Netherlands
FEATURES
Location/Qualifiers
1..3923
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
```

```
repeat_region /map="9q21-q22"
3756..3911 /tissue_type="prostate"
/note="non-coding RNA DD3"
/note="LINE"
/rpt_family="Li"
/rpt_type="dispersed"
polya_site 3923
BASE COUNT 1157 a 840 c 740 g 1186 t
ORIGIN
Query Match 88.5%; Score 725.4; DB 88; Length 3923;
Best Local Similarity 99.9%; Pred. No. 5.8e-195;
Matches 726; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 94 cagaggtgagaaataagaaggctgtgactttacacatctgagccacacatctgtctgaa 153
Db 281 CAGGGGTGAGAAATAGAAAGGCTGTGACTTTTACCATCTGAGGCCACACATCTGCTGAA 340
QY 154 atggagataatataacatcactagaaacagcaagatgacaataataatgtctaaagtac 213
Db 341 ATGGAGATATTAACATCACTAGAAACAGCAAGATGACAATAATATGTCTAAGTAGTGAC 400
QY 214 atgttttgcacatttccagccctttaaatatccacacacaggaagcacaaaggaa 273
Db 401 ATGTTTTGCACATTTCCAGCCCTTTAAATATCCACACACAGGAAGCACAAGAGGAA 460
QY 274 gcacagagatccctggggagaaatgccggccgcctctgtggtcacatcgatgagcctgc 333
Db 461 GCACAGAGATCCCTGGGAGAAATGCCGGCCGCATCTTTGGGTCATCGATGAGCCTCGCC 520
QY 334 ctgtgctgtctccgcttctgagggaagacattagaaaaatgaatgattgttctcttaa 393
Db 521 CHTGTGCTGGTCCCGCTTGTGAGGGAAGACATTAGAAAATGAATGTGTGTCTCTTAA 580
QY 394 aggatgggagaaacagacatcctgtgtggatatttatttgacgggattacagattg 453
Db 581 AGGATGGCAGGAGAAACAGATCCTGTTGTGGATATTTATTTGAACGGGATTACAGATTG 640
QY 454 aaatgaagtcacaaagtgcattaccatgagagagaaacagagagaaatcttgatg 513
Db 641 AAATGAAGTCACAAAGTGAGCATTACCAATGAGAGGAGAAACAGAGAGAAATCTTTGATG 700
QY 514 gcttcacagacatgcacaaacaaatactgtgatactgtgatgacatgagggcagccaagct 573
Db 701 GCTTCACAGACATGCACAAACAAATGAATGATCTGTGATGATGATGAGGCAGCAGCT 760
QY 574 ggggaggagataaccacggggcagagggtcaggattcttgccctgctgctaaactgtgc 633
Db 761 GGGGAGGAGATAACACACGGGCGAGAGGGTCAGGATTTCTGGCCCTGCTCTAAACTGTC 820
QY 634 gttcataacaaataatcttcatatttctaacccctcaaaaaaagctgttgaataatctga 693
Db 821 GTTCATAACCAAAATCATATTTTCATATTTTAACTCAAAACAAAGCTGTGTGTAATATCTGA 880
QY 694 tctctacgattcttctggcccaacatctccatatatccacccacacatcttttaatt 753
Db 881 TCTCTACGGTTCCTTCTGGGCCCAACATCTCTCCATATATCCACCCACACTCATTTTTAAT 940
QY 754 attagtctccagatctgtactgtgacctttctacactgtagataacattactctttt 813
Db 941 ATTAGTCTCCAGATCTGTACTGTGACCTTTCTACACTGTAGATAACATTACTCATTTT 1000
QY 814 gttcaaaa 820
Db 1001 GTTCAAA 1007
RESULT 2
AF103908
LOCUS AF103908 5435 bp DNA PRI 14-AUG-2000
DEFINITION Homo sapiens non-coding RNA DD3 gene, exons 2, 3, and 4.
```


ACCESSION AF103908
VERSION AF103908.1 GI:6165974
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 5435)
AUTHORS Bussemakers, M.J., van Bokhoven, A., Verhaegh, G.W., Smit, F.P.,
Karthaas, H.F., Schalken, J.A., Debruyne, F.M., Ru, N. and Isaacs, W.B.
TITLE DD3: a new prostate-specific gene, highly overexpressed in prostate
cancer
JOURNAL Cancer Res. 59 (23), 5975-5979 (1999)
MEDLINE 20072260
PUBMED 10606244
REFERENCE 2 (bases 1 to 5435)
AUTHORS Bussemakers, M.J.G., Van Bokhoven, A., Verhaegh, G.W., Smit, F.P.,
Karthaas, H.F.M., Schalken, J.A., Debruyne, F.M.J., Ru, N. and
Isaacs, W.B.
TITLE Direct Submission
JOURNAL Submitted (28-OCT-1998) Urology Research Laboratory, University
Hospital Nijmegen, P.O. Box 9101, Nijmegen 6500 HB, Netherlands
FEATURES
source
1..5435
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/map="9q21-q22"
1..78
/rpt_family="Alu"
/rpt_type="dispersed"
533..697
/note="alternative exon present in 5% of cDNA clones"
/number=2
repeat_region 1035..1294
/rpt_family="Alu"
/rpt_type="dispersed"
join(1571..1753,1981..5435)
/note="transcript I1"
/product="non-coding RNA DD3"
join(1571..1753,1981..2517)
/note="transcript I"
/product="non-coding RNA DD3"
join(1571..1753,1981..3579)
/note="transcript (major) II"
/product="non-coding RNA DD3"
1571..1753
/number=3
exon 1981..5435
/number=4
polyA_signal 2495..2499
/note="transcript I"
polyA_site 2517
/note="transcript I"
polyA_signal 3553..3558
/note="transcript (major) II"
polyA_site 3579
/note="transcript (major) II"
repeat_region 5268..5423
/note="LINE"
/rpt_family="L1"
/rpt_type="dispersed"
polyA_site 5435
/note="transcript III"
BASE COUNT 1546 a 1188 c 1036 g 1654 t 11 others
ORIGIN

Query Match 66.1%; Score 542.4; DB 88; Length 5435;
Best Local Similarity 99.8%; Pred. No. 5.4e-143;
Matches 543; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 277 cagagatccctgggagaaatgccgcgcgcctcttgggtcatcgatgagcctgcgcctg 336

Db 1976 CACAGATCCCTGGGAGAAATGCCGCCCATCTTGGGTCTCATGATGAGCCCTGCGCCTG 2035
Qy 337 tgcctggtcccgctgtgaggaagacattagaaataaattgattgtgtcttcttaaaag 396
Db 2036 TGCCCTGGTCCCGCTGTGTAGGGAAGACATTAGAAAATGAATGTGATGTCTTCTTAAGG 2095
Qy 397 atggcagagaaacagatcctgttggatatttttgaacgggattacagatttgaag 456
Db 2096 ATGGGCAGGAAAACAGATCCTCTGTGGATATTATTTCACGGGATACAGATTTGAAA 2155
Qy 457 tgaagtacaaagttagcattaccatgagagagagaaacagacagagaaaaatcttgatg 516
Db 2156 TGAAGTCAAAAAGTGAGCAATTACCAATGAGAGAAAACAGACAGAAAATCTTGATGGCT 2215
Qy 517 tcacaagacatcaacaaacaaatggaatactgtgatacatgagcagcacaactgag 576
Db 2216 TCACAGAGATCAACAAACAAATGGAATCTGTGATGACATGAGGACCAAGCTGGG 2275
Qy 577 gaggagataaccacggggcagagggctcagattctggccctgctgcctaaactgtgcgt 636
Db 2276 GAGGAGATAACCAACGCGGAGAGGCTCAGGATTCGTGCCCTGCTAACTGTGCGTT 2335
Qy 637 cataccaaatcatttcatttctaaacctcaaaacaaagctgtgtgataatctgatct 696
Db 2336 CATAACCAATCATTTTCATATTCTAAACCTCAAAACAAAGCTGTGTAATATCTGATCT 2395
Qy 697 ctacggtctctctggggcccaacttccatataatccacacacactcttttaatt 756
Db 2396 CTACGGTTCCTTCTGGGCCCAACATTCATATATATCCAGCCACACTCAATTTTAATAT 2455
Qy 757 tagtccacagatctgactgtgacctttctacactgtagaataaacattactattt 816
Db 2456 TAGTTCACAGATCTGACTGTGACCTTCTACACTGTAGATAACATTACTATTGTT 2515
Qy 817 caaa 820
Db 2516 CAAA 2519

RESULT 3
AL390239
LOCUS Homo sapiens chromosome 9 clone RP11-58J3, *** SEQUENCING IN
DEFINITION PROGRESS ***, 4 unordered pieces.
ACCESSION AL390239
VERSION AL390239.11 GI:13274794
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 164371)
AUTHORS Burton, J.
TITLE Direct Submission
JOURNAL Submitted (05-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT On Mar 12, 2001 this sequence version replaced gi:13273805.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA58J3
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 16337 bases at least Q40
Consensus quality: 163481 bases at least Q30
Consensus quality: 163577 bases at least Q20

Insert size: 164071; sum-of-contigs
 Insert size: 166918; 4.2% error; agarose-fp
 Quality coverage: 8.78x in Q20 bases; sum-of-contigs Quality
 coverage: 9.07x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 20240: contig of 20240 bp in length
 * 20241 20340: gap of 100 bp
 * 20341 116100: contig of 95760 bp in length
 * 116101 116200: gap of 100 bp
 * 116201 118243: contig of 2043 bp in length
 * 118244 118343: gap of 100 bp
 * 118344 164371: contig of 46028 bp in length.
 Location/Qualifiers
 1..164371
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="9"
 /clone="RP11-5873"
 /clone_lib="RPC1-11.1"
 1..20240
 /note="assembly_fragment:00838
 fragment_chain:1
 clone_end:SP6
 vector_side:left"
 20341..116100
 /note="assembly_fragment:01998
 fragment_chain:1"
 116201..118243
 /note="assembly_fragment:00916
 fragment_chain:1"
 118344..164371
 /note="assembly_fragment:00334
 fragment_chain:1
 clone_end:T7
 vector_side:right"
 300 others
 BASE COUNT 50254 a 34905 c 32663 g 46249 t
 ORIGIN
 Query Match 66.1%; Score 542.4; DB 80; Length 164371;
 Best Local Similarity 99.8%; Pred. No. 9.1e-143;
 Matches 543; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 277 cagagatccctggagaaatgccggccgcatcttgggtcatcgatgagcctgcctg 336
 |||
 Db 33316 CACAGATCCCTGGGAGAAATGCCGCCGCCATCTTGGGTCTCGATGAGCCCTG 33375
 |||
 QY 337 tgctgtccctgtgtgaggaagacattagaaatgaattgattgttccctaaag 396
 |||
 Db 33376 TGCTGTCCCGCTGTGAGGGAAGGACATTAGAAATGAATTGATGTCTCTTAAGG 33435
 |||
 QY 397 atggcaggagaaacagatcctgtgtggtattatttgaacgggattacagattgaa 456
 |||
 Db 33436 ATGGCAGGAGAAACAGATCCTGTGTGGATATTTATTGAACGGGATTACAGATTGAAA 33495
 |||
 QY 457 tgaagtcaacaagtgaattacattacagagaaacacagagagaaatcttgatgct 516
 |||
 Db 33496 TGAAGTCAACAAGTGAAGCAATACCAATGAGAGAAACACAGAGAAATCTTGATGGCT 33555
 |||
 QY 517 tcacaagacatgcacaacaaatggaattactgtgatgacatgagcgagcagctggg 576
 |||
 Db 33556 TCACAAGACATGCACAACAAATGGAATACGTGATGACATGAGCAGCAGCTGGG 33615
 |||
 QY 577 gaggagataaccacggggcagaggggtcaggattctggcctgctgctaaactgtcgctt 636
 |||

Db 33616 GAGGAGATAACCCAGGGGCGAGGGGTGAGGATTTCTGCCCTGCTGCTAAACTGTCGTT 33675
 QY 637 cataacccaatcatcttcatatttcaacctcaaaacaaagctgtgtgtaattctgatct 696
 |||
 Db 33676 CATACCAATCATTTTCATATTTCTAACCTCAAAACAAGCTGTGTATATATCTGATCT 33735
 QY 697 ctacggttctcttgggcccaacattctccatatatccagccacacactcatttttaatt 756
 |||
 Db 33736 CTACGGTCTCTTCTGGGCCCAACATTTCTCATATATCCAGCCACACTCATTTTAATAT 33795
 QY 757 tagttcccgatctgtactgtgacctttctacactgtagaataacattactatttgg 816
 |||
 Db 33796 TAGTCCAGATCTGTACTGTGACCTTTCTACACTGTAGATAACATTTACTCATTTTGT 33855
 QY 817 caaa 820
 |||
 Db 33856 CAAA 33859
 |||
 RESULT 4
 AL359314
 LOCUS AL359314 173831 bp DNA HTG 08-APR-2001
 DEFINITION Homo sapiens chromosome 9 clone RP11-10814, *** SEQUENCING IN
 PROGRESS ***, 3 unordered pieces.
 ACCESSION AL359314
 VERSION AL359314.12 GI:13396560
 KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 173831)
 Leongamornlert, D.
 Direct Submission
 Submitted (07-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Mar 20, 2001 this sequence version replaced gi:13277120.
 ----- Genome Center
 Center: Sanger Centre
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquery@sanger.ac.uk
 ----- Project Information
 Center project name: BA10814
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 173545 bases at least Q40
 Consensus quality: 173599 bases at least Q30
 Consensus quality: 173621 bases at least Q20
 Insert size: 173631; sum-of-contigs
 Insert size: 172123; 10.0% error; agarose-fp
 Quality coverage: 10.35x in Q20 bases; sum-of-contigs Quality
 coverage: 10.50x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 50595: contig of 50595 bp in length
 * 50596 50695: gap of 100 bp
 * 50696 157636: contig of 106941 bp in length
 * 157637 157736: gap of 100 bp
 * 157737 173831: contig of 16095 bp in length.
 Location/Qualifiers
 1..173831
 /organism="Homo sapiens"
 FEATURES
 source

db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-108L4"
/clone_lib="RPC1-11.1"
1..50595
/note="assembly_fragment:04137
fragment_chain:1"
50696..157636
/note="assembly_fragment:04236
fragment_chain:1"
157737..173831
/note="assembly_fragment:01656"
BASE COUNT 52107 a 36469 c 35018 g 50037 t 200 others
ORIGIN

Query Match 66.1%; Score 542.4; DB 80; Length 173831;
Best Local Similarity 99.8%; Pred. No. 9.2e-143;
Matches 543; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 277 cagagatccctggagaaatcccgccgcctcttgggtcatcgatgagcctcgccctg 336
|||
Db 156237 CACAGATCCCTGGGAGAAATCCCGCCGCATCTTGGGTCTATGATGAGCCTCGCCCTG 156296
QY 337 tgctggtccctgtgtgagggaagacattagaaatgaattgatgtgttccctaaagg 396
|||
Db 156297 TGCTGTGTCCTGTTGTGAGGGAAGGACATTAGAAATGAATTGATGTCTCTTTAAAGG 156356
QY 397 atgggcagagaaacagatcctgtgtggtatttatttgaacgggttacagatttgaaa 456
|||
Db 156357 ATGGCGAGGAAACAGATCCTGTTGGGATATTTATTTGAACGGGATTACAGATTTGAAA 156416
QY 457 tgaagtccaaagtgcattaccatgagagaaacagacagagaaatcttgatgctt 516
|||
Db 156417 TGAAGTCACAAAGTAGCATTTACCAATGAGAGGAAACAGACGAGAAATCTTGATGGCT 156476
QY 517 tcaagagacatgcaacaaacaaatggaatactgtgatgacatgagcgagcgaagctggg 576
|||
Db 156477 TCACAGAGATGCAACAAACAAATGGAATGATGTGATGACATGAGGCGCAAGCTGGG 156536
QY 577 gaggagataaccaggggagagggcagagattctggtcctgctgctgaactgtgcgtt 636
|||
Db 156537 GAGGAGATAACCAACGGGGCAGAGGGTCAGGATCTCGCCCTGCTGCTAAACTGTCGTT 156596
QY 637 cataacaaatcatttcatttaccctcaaccccaaaacgaagctgtgtgaatctgatct 696
|||
Db 156597 CATACCAATCATTTTCATTTCTAACCTCAAAACAAGCTGTGTATATCTCATCT 156656
QY 697 ctacggttcccttggtggcccaacattctccatatatccagccacactcatcttttaatt 756
|||
Db 156657 CTACGGTTCTCTGCGGCCCAACATTTCTCATATATCCAGCCACACTCATTTTAAATAT 156716
QY 757 taagtccagatctgtactgtacacctttacactgtagaataaacattactcaatttgg 816
|||
Db 156717 TAGTTCAGATCTGTACTGTGACCTTTTACACTGTAGAATAACATTACTCATTTTGT 156776
QY 817 caaa 820
|||
Db 156777 CAAA 156780

RESULT 5
AL358573/c
LOCUS
DEFINITION Homo sapiens chromosome 9 clone RP11-133022, *** SEQUENCING IN
PROGRESS ***, 37 unordered pieces.
ACCESSION AL358573
VERSION AL358573.17 GI:13660951
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 267581)
Mclay, K.
Direct Submission
Submitted (14-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequests@sanger.ac.uk
On Apr 17, 2001 this sequence version replaced gi:13398774.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bal33022
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 256699 bases at least Q40
Consensus quality: 259744 bases at least Q30
Consensus quality: 261407 bases at least Q20
Insert size: 263981; sum-of-contigs
Insert size: 135491; 19.3% error; agarose-fp
Quality coverage: 5.51x in Q20 bases; sum-of-contigs Quality
coverage: 12.23x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 25718: contig of 25718 bp in length
* 25719 25818: gap of 100 bp
* 25819 29567: contig of 3749 bp in length
* 29568 29667: gap of 100 bp
* 29668 40520: contig of 10853 bp in length
* 40521 40620: gap of 100 bp
* 40621 50002: contig of 9382 bp in length
* 50003 50102: gap of 100 bp
* 50103 54942: contig of 4840 bp in length
* 54943 55042: gap of 100 bp
* 55043 59067: contig of 4025 bp in length
* 59068 59167: gap of 100 bp
* 59168 72327: contig of 13160 bp in length
* 72328 72427: gap of 100 bp
* 72428 79396: contig of 6969 bp in length
* 79397 79496: gap of 100 bp
* 79497 82738: contig of 3242 bp in length
* 82739 82838: gap of 100 bp
* 82839 87543: contig of 4705 bp in length
* 87544 87643: gap of 100 bp
* 87644 89847: contig of 2204 bp in length
* 89848 89947: gap of 100 bp
* 89948 92376: contig of 2429 bp in length
* 92377 92476: gap of 100 bp
* 92477 99799: contig of 7323 bp in length
* 99800 99899: gap of 100 bp
* 99900 102117: contig of 2218 bp in length
* 102118 102217: gap of 100 bp
* 102218 109330: contig of 7113 bp in length
* 109331 109430: gap of 100 bp
* 109431 112187: contig of 2757 bp in length
* 112188 112287: gap of 100 bp
* 112288 116407: contig of 4120 bp in length
* 116408 116507: gap of 100 bp
* 116508 119386: contig of 2879 bp in length
* 119387 119486: gap of 100 bp
* 119487 121869: contig of 2383 bp in length
* 121870 121969: gap of 100 bp
* 121970 125546: contig of 3577 bp in length

* 125547 125646: gap of 100 bp
* 125647 128990: contig of 3344 bp in length
* 128991 129090: gap of 100 bp
* 129091 134055: contig of 4965 bp in length
* 134056 134155: gap of 100 bp
* 134156 138314: contig of 4159 bp in length
* 138315 138414: gap of 100 bp
* 138415 140612: contig of 2198 bp in length
* 140613 140712: gap of 100 bp
* 140713 143296: contig of 2584 bp in length
* 143297 143396: gap of 100 bp
* 143397 148113: contig of 4717 bp in length
* 148114 148213: gap of 100 bp
* 148214 151009: contig of 2796 bp in length
* 151010 151109: gap of 100 bp
* 151110 156616: contig of 5507 bp in length
* 156617 156716: gap of 100 bp
* 156717 158763: contig of 2047 bp in length
* 158764 158863: gap of 100 bp
* 158864 162129: contig of 3266 bp in length
* 162130 162229: gap of 100 bp
* 162230 164443: contig of 2214 bp in length
* 164444 164543: gap of 100 bp
* 164544 169800: contig of 5257 bp in length
* 169801 169900: gap of 100 bp
* 169901 251545: contig of 81645 bp in length
* 251546 251645: gap of 100 bp
* 251646 259029: contig of 7384 bp in length
* 259030 259129: gap of 100 bp
* 259130 262871: contig of 3742 bp in length
* 262872 262971: gap of 100 bp
* 262972 265209: contig of 2238 bp in length
* 265210 265309: gap of 100 bp
* 265310 267581: contig of 2272 bp in length.

FEATURES

Source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-133022"
/clone_lib="RP11-11.1"
misc_feature
1..25718
/note="assembly_fragment:00645
fragment_chain:1
clone_end:T7
vector_side:left"
misc_feature
25819..29567
/note="assembly_fragment:01660
fragment_chain:1"
misc_feature
29668..40520
/note="assembly_fragment:03080
fragment_chain:1"
misc_feature
40621..50002
/note="assembly_fragment:00771
fragment_chain:1"
misc_feature
50103..54942
/note="assembly_fragment:02629
fragment_chain:1"
misc_feature
55043..59067
/note="assembly_fragment:04565
fragment_chain:1"
misc_feature
59168..72327
/note="assembly_fragment:00223
fragment_chain:1"
misc_feature
72428..79396
/note="assembly_fragment:03318
fragment_chain:1"
misc_feature
79497..82738
/note="assembly_fragment:00224
fragment_chain:2"
misc_feature
82839..87543
/note="assembly_fragment:03050
fragment_chain:2"

misc_feature
87644..89847
/note="assembly_fragment:01948
fragment_chain:3"
misc_feature
89948..92376
/note="assembly_fragment:02191
fragment_chain:3"
misc_feature
92477..99799
/note="assembly_fragment:00010"
misc_feature
99900..102117
/note="assembly_fragment:00672"
misc_feature
102218..109330
/note="assembly_fragment:01326"
misc_feature
109431..112187
/note="assembly_fragment:01498"
misc_feature
112288..116407
/note="assembly_fragment:01607"
misc_feature
116508..119386
/note="assembly_fragment:01899"
misc_feature
119487..121869
/note="assembly_fragment:02068"
misc_feature
121970..125546
/note="assembly_fragment:02469"
misc_feature
125647..128990
/note="assembly_fragment:02502"
misc_feature
129091..134055
/note="assembly_fragment:02669"
misc_feature
134156..138314
/note="assembly_fragment:02757"
misc_feature
138415..140612
/note="assembly_fragment:02776"
misc_feature
140713..143296
/note="assembly_fragment:03127"
misc_feature
143397..148113
/note="assembly_fragment:03634"
misc_feature
148214..151009
/note="assembly_fragment:03898"
misc_feature
151110..156616
/note="assembly_fragment:04220"
misc_feature
156717..158763
/note="assembly_fragment:04276"
misc_feature
158864..162129
/note="assembly_fragment:04418"
misc_feature
162230..164443
/note="assembly_fragment:04428"

Query Match 66.1%; Score 542.4; DB 80; Length 267581;
Best Local Similarity 99.8%; Pred.No.9.8e-143;
Matches 543; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 277 cagagatccctgggagaaatgccggccgccatcttgggtcatcgatgagcctcgccctg 336
|| |||||
Db 165900 CACAGATCCCTGGGAGAAATGCCGCCGCCCATCTTGGGTCTATCGATGAGCTCGCCCTG 165841
QY 337 tgcctggctccctgtgtgaggaagacattagaaaaatgaattgatgtgtccttaagg 396
|||||
Db 165840 TGCCTGGTCCCGCTTGTGAGGAAGACATTAGAAAAATGAATTGATGTCTTCTTAAAGG 165781
QY 397 atggcagagaaacagatcctgttggatatttttgaacggattacacatttga 456
|||||
Db 165780 ATGGGCAGGAAACAGATCCTGTGTGGATATTTATTTGAACGGGATTACAGATTGTGAAA 165721
QY 457 tgaagtcaacaagtgagcattaccatgagaggaacacagagagaaaaatccttgagct 516
|||||
Db 165720 TGAAGTCACAAAGTGAGCATACCAATGAGAGGAAACACAGAGAAAAATCTTGATGGCT 165661
QY 517 tcacaagacatgcaacaacaaatggaatactgtgatacagatgagggcagcaagctgg 576
|||||
Db 165660 TCACAAGACATGCAACAAACAAATGGAATACTGTGATGACATGAGGACCAAGCTGGG 165601
QY 577 gaggagataacacaggggcagaggttcaggattctggccctcgtcgtcctaaactgtcgtt 636
|||||
Db 165600 GAGGAGATTAACACGGGGCAGAGGGTTCAGGATTCCTGCGCTTAACACTGTCGTT 165541

```
QY 637 catacaaaatcatttcatttctaaacctcaaaagaagctgttgtaatactatctgatct 696
|||||
Db 165540 CATAACCAATCATTTTCATATTTCTAAACCTCAAAACAAGCTGTGTATATCTGATCT 165481
|||||
QY 697 ctacggctcctctggggccacattctccatatatccagccacactcatttttaattatt 756
|||||
Db 165480 CTACGGTCTCTTCTGGGCGCAACATTTCTCATATATCCAGCCACACTCATTTTAAATAT 165421
|||||
QY 757 tagttccagatctgactgtgacctttctacactgtagaataaacattactcattttgtt 816
|||||
Db 165420 TAGTTCACAGATCTGACTGTGACCTTTCTACACTGTAGATAACATTACTCATTTTGT 165361
|||||
QY 817 caaa 820
|||||
Db 165360 CAAA 165357

RESULT 6
AX018075 359 bp DNA PAT 07-SEP-2000
LOCUS Sequence 23 from Patent W09946374.
DEFINITION AX018075
ACCESSION AX018075
VERSION AX018075.1 GI:10042526
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 359)
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Pilarczyk,C.
TITLE Human nucleic acid sequences from prostate tumour tissue
JOURNAL Patent: WO 9946374-A 23 16-SEP-1999;
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILARSKY CHRISTIAN (DE)
FEATURES
    source
        1..359
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
BASE COUNT 121 a 75 c 94 g 69 t
ORIGIN

Query Match 40.2%; Score 330; DB 9; Length 359;
Best Local Similarity 100.0%; Pred. No. 7.1e-83;
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agaagctggcatcagaaaaacagagggagatttgtgtgctgcagccgagggagaccag 60
|||||
Db 30 AGAAGCTGGCATCAGAAAAACAGAGGGGAGATTGTGTGCTGCAGCCCGAGGAGACCAG 89
|||||
QY 61 gaagatctgcattgtgggaagacctgtatgatacagaggtgagaaaataagaaaggctgt 120
|||||
Db 90 GAAGATCTGCATTGTGGGAAGGACCTGTATGATACAGAGGTGAGAAAATAAGAAAGGCTGT 149
|||||
QY 121 gactttaccatctgagccacacatctctgaaatggagataataaactaactagaacaa 180
|||||
Db 150 GACTTTACCATCTGAGGCCACACATCTGCTGAATGGAGATRAATTAACTACTAGTAAC 209
|||||
QY 181 agcaagatgacaataatgtctaaagttagtgacatgtttttgcacattttccagcccttt 240
|||||
Db 210 AGCAAGATGACAATAATGTCTAAGTAGTACATGTTTTCACATTTCCAGCCCTTT 269
|||||
QY 241 aaatattccacacacagagaagcacaaaaaggaagcacagagatccctgggagaaaatgc 300
|||||
Db 270 AAATATCCACACACAGGAAGCACAAAAGGAAGCACAGAGATCCCTGGGAGAAATGCC 329
|||||
QY 301 gggccgcatcttggtcatgatgagcctc 330
|||||
Db 330 GGCCCGCATCTTGGGTGTCATCGATGAGCCCTC 359
|||||
```

```
RESULT 7
AL161625 143675 bp DNA HTG 20-JAN-2001
LOCUS Homo sapiens chromosome 9 clone RP11-146P9, *** SEQUENCING IN
DEFINITION PROGRESS ***, 10 unordered pieces.
ACCESSION AL161625
VERSION AL161625.6 GI:9863607
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 143675)
AUTHORS Plumb,B.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 21, 2000 this sequence version replaced gi:8894260.
COMMENT
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bA146P9
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 138647 bases at least Q40
Consensus quality: 140619 bases at least Q30
Consensus quality: 141706 bases at least Q20
Insert size: 142775; sum-of-contigs
Insert size: 147523; 5.9% error; agarose-fp
Quality coverage: 4.29x in Q20 bases; sum-of-contigs Quality
coverage: 4.22x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 12163: contig of 12163 bp in length
* 12164 12263: gap of 100 bp
* 12264 25269: contig of 13006 bp in length
* 25270 25369: gap of 100 bp
* 25370 57709: contig of 32340 bp in length
* 57710 57809: gap of 100 bp
* 57810 66792: contig of 8983 bp in length
* 66893 102772: contig of 35880 bp in length
* 102773 102872: gap of 100 bp
* 102873 106863: contig of 3991 bp in length
* 106864 106963: gap of 100 bp
* 106964 129804: contig of 22841 bp in length
* 129805 129904: gap of 100 bp
* 129905 132223: contig of 2319 bp in length
* 132224 132323: gap of 100 bp
* 132324 135764: contig of 3441 bp in length
* 135765 135864: gap of 100 bp
* 135865 143675: contig of 7811 bp in length.
FEATURES
    Location/Qualifiers
        1..143675
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="9"
            /clone="RP11-146P9"
            /clone_lib="RPCI-11.1"
        misc_feature
            1..12163
```

```

/note="assembly_fragment:00440
clone_end:77
vector_side:left"
misc_feature
12264..25269
/note="assembly_fragment:00434
fragment_chain:1"
misc_feature
25370..57709
/note="assembly_fragment:00526
fragment_chain:1"
misc_feature
57810..66792
/note="assembly_fragment:00661
fragment_chain:1"
misc_feature
66893..102772
/note="assembly_fragment:00323"
102873..106863
/note="assembly_fragment:00860"
106964..129804
/note="assembly_fragment:00978"
129905..132223
/note="assembly_fragment:01068"
132324..135764
/note="assembly_fragment:01424"
135865..143675
/note="assembly_fragment:01413
clone_end:SP6
vector_side:right"
BASE COUNT 42329 a 30042 c 28964 g 41431 t 909 others
ORIGIN

```

```

Query Match      12.2%; Score 100.2; DB 79; Length 143675;
Best Local Similarity 97.1%; Pred. No. 3e-17;
Matches 102; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 agaagctgcacagaaaacagagggagattgtgtgctgcacccagggagacag 60
Db 136065 AGAAGCTGGCTCAGAAAACAGAGGGGAGATTGTGTGCTGCACCCGAGGAGACCAG 136124

QY 61 gaagatctgcattgtgggaagacctgatcatcacagaggtgagaa 105
Db 136125 GAAGATCTGCATGGTGAAGGACCTGATCATCACAGAGGTCTGTA 136169

```

```

RESULT 8
AL358573
LOCUS
DEFINITION Homo sapiens chromosome 9 clone RP11-133022, HTG 15-APR-2001
PROGRESS ***, 37 unordered pieces.
ACCESSION AL358573
VERSION AL358573.17 GI:13660951
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 267581)
Direct Submission
Submitted (14-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Apr 17, 2001 this sequence version replaced gi:13398774.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bal33022
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads

```

```

Consensus quality: 256699 bases at least Q40
Consensus quality: 259744 bases at least Q30
Consensus quality: 261407 bases at least Q20
Insert size: 263981; sum-of-contigs
Insert size: 135491; 19.3% error; agarose-fp
Quality coverage: 5.51x in Q20 bases; sum-of-contigs Quality
coverage: 12.23x in Q20 bases; agarose-fp
-----

```

* NOTE: This is a 'working draft' sequence. It currently consists of 37 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 25718: contig of 25718 bp in length
* 25719 25818: gap of 100 bp
* 25819 29567: contig of 3749 bp in length
* 29568 29667: gap of 100 bp
* 29668 40520: contig of 10853 bp in length
* 40521 40620: gap of 100 bp
* 40621 50002: contig of 9382 bp in length
* 50003 50102: gap of 100 bp
* 50103 54942: contig of 4840 bp in length
* 54943 55042: gap of 100 bp
* 55043 59067: contig of 4025 bp in length
* 59068 59167: gap of 100 bp
* 59168 72327: contig of 13160 bp in length
* 72328 72427: gap of 100 bp
* 72428 79396: contig of 6969 bp in length
* 79397 79496: gap of 100 bp
* 79497 82738: contig of 3242 bp in length
* 82739 82838: gap of 100 bp
* 82839 87543: contig of 4705 bp in length
* 87544 87643: gap of 100 bp
* 87644 89847: contig of 2204 bp in length
* 89848 89947: gap of 100 bp
* 89948 92376: contig of 2429 bp in length
* 92377 92476: gap of 100 bp
* 92477 99799: contig of 7323 bp in length
* 99800 99899: gap of 100 bp
* 99900 102117: contig of 2218 bp in length
* 102118 102217: gap of 100 bp
* 102218 109330: contig of 7113 bp in length
* 109331 109430: gap of 100 bp
* 109431 112187: contig of 2757 bp in length
* 112188 112287: gap of 100 bp
* 112288 116407: contig of 4120 bp in length
* 116408 116507: gap of 100 bp
* 116508 119386: contig of 2879 bp in length
* 119387 119486: gap of 100 bp
* 119487 121869: contig of 2383 bp in length
* 121870 121969: gap of 100 bp
* 121970 125546: contig of 3577 bp in length
* 125547 125646: gap of 100 bp
* 125647 128990: contig of 3344 bp in length
* 128991 129090: gap of 100 bp
* 129091 134055: contig of 4965 bp in length
* 134056 134155: gap of 100 bp
* 134156 138314: contig of 4159 bp in length
* 138315 138414: gap of 100 bp
* 138415 140612: contig of 2198 bp in length
* 140613 140712: gap of 100 bp
* 140713 143296: contig of 2584 bp in length
* 143297 143396: gap of 100 bp
* 143397 148113: contig of 4717 bp in length
* 148114 148213: gap of 100 bp
* 148214 151009: contig of 2796 bp in length
* 151010 151109: gap of 100 bp
* 151110 156616: contig of 5507 bp in length
* 156617 156716: gap of 100 bp
* 156717 158763: contig of 2047 bp in length

```

```

* 158764 158863: gap of 100 bp
* 158864 162129: contig of 3266 bp in length
* 162130 162229: gap of 100 bp
* 162230 164443: contig of 2214 bp in length
* 164444 164543: gap of 100 bp
* 164544 169800: contig of 5257 bp in length
* 169801 169900: gap of 100 bp
* 169901 251545: contig of 81645 bp in length
* 251546 251645: gap of 100 bp
* 251646 259029: contig of 7384 bp in length
* 259030 259129: gap of 100 bp
* 259130 262871: contig of 3742 bp in length
* 262872 262971: gap of 100 bp
* 262972 265209: contig of 2238 bp in length
* 265210 265309: gap of 100 bp
* 265310 267581: contig of 2272 bp in length.
FEATURES
    source
        1. .267581
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="9"
            /clone="RP11-133022"
            /clone_lib="RPC1-II.1"
        1. .25718
            /note="assembly_fragment:00645"
            fragment_chain:1
            clone_end:T7
            vector_side:left"
        25819. .29567
            /note="assembly_fragment:01660"
            fragment_chain:1"
        29668. .40520
            /note="assembly_fragment:03080"
            fragment_chain:1"
        40621. .50002
            /note="assembly_fragment:00771"
            fragment_chain:1"
        50103. .54942
            /note="assembly_fragment:02629"
            fragment_chain:1"
        55043. .59067
            /note="assembly_fragment:04565"
            fragment_chain:1"
        59168. .72327
            /note="assembly_fragment:00223"
            fragment_chain:1"
        72428. .79396
            /note="assembly_fragment:03318"
            fragment_chain:1"
        79497. .82738
            /note="assembly_fragment:00224"
            fragment_chain:2"
        82839. .87543
            /note="assembly_fragment:03050"
            fragment_chain:2"
        87644. .89847
            /note="assembly_fragment:01948"
            fragment_chain:3"
        89948. .92376
            /note="assembly_fragment:02191"
            fragment_chain:3"
        92477. .99799
            /note="assembly_fragment:00010"
            fragment_chain:3"
        99900. .102117
            /note="assembly_fragment:00672"
            fragment_chain:3"
        102218. .109330
            /note="assembly_fragment:01326"
            fragment_chain:3"
        109431. .112187
            /note="assembly_fragment:01498"
            fragment_chain:3"
        112288. .116407
            /note="assembly_fragment:01607"
            fragment_chain:3"
        116508. .119386
            /note="assembly_fragment:01899"
            fragment_chain:3"

```

```

misc_feature 119487. .121869
              /note="assembly_fragment:02068"
misc_feature 121970. .125546
              /note="assembly_fragment:02469"
misc_feature 125647. .128990
              /note="assembly_fragment:02502"
misc_feature 129091. .134055
              /note="assembly_fragment:02669"
misc_feature 134156. .138314
              /note="assembly_fragment:02757"
misc_feature 138415. .140612
              /note="assembly_fragment:02776"
misc_feature 140713. .143296
              /note="assembly_fragment:03127"
misc_feature 143397. .148113
              /note="assembly_fragment:03634"
misc_feature 148214. .151009
              /note="assembly_fragment:03898"
misc_feature 151110. .156616
              /note="assembly_fragment:04220"
misc_feature 156717. .158763
              /note="assembly_fragment:04276"
misc_feature 158864. .162129
              /note="assembly_fragment:04418"
misc_feature 162230. .164443
              /note="assembly_fragment:04428"

Query Match 12.2%; Score 100.2; DB 80; Length 267581;
Best Local Similarity 97.1%; Pred. No. 3.3e-17;
Matches 102; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 agaagctgcgcacagaaaacagaggagatttgtgtggtgcagcgagagagaccag 60
    |||||
Db 253275 AGAAGCTGCATCAGAAAAACAGAGGGAGGAGATTGTGTGGTGCAGCGGAGACCAG 253334
    |||||

Qy 61 gaagatctgcattgtgggaagacctgatgatacagaggtgagaa 105
    |||||
Db 253335 GAAGATCTGCATGTTGGGAAGGACCTGATGATACAGAGGTCTGTA 253379
    |||||

RESULT 9
AC009556/c 172298 bp DNA HTG 26-MAY-2000
LOCUS Homo sapiens clone RP11-57C21, WORKING DRAFT SEQUENCE, 26 unordered
DEFINITION pieces.
ACCESSION AC009556
VERSION AC009556.4 GI:7107752
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 172298)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
TITLE Karas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
JOURNAL Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
REFERENCE Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
AUTHORS Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
            Direct Submission
TITLE

```

JOURNAL

COMMENT

Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 28, 2000 this sequence version replaced gi:6479158.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RN/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1672
Center clone name: 57_C_21
----- Summary Statistics

Sequencing vector: M13: M7815; 100% of reads
Chemistry: Dye-terminator-amersham; 5% of reads
Assembly: Dye-terminator Big Dye; 95% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 136385 bases at least Q40
Consensus quality: 156543 bases at least Q30
Consensus quality: 165534 bases at least Q20
Insert size: 165000; agarose-fp
Insert size: 169798; sum-of-contigs
Quality coverage: 4.2 in Q20 bases; agarose-fp
Quality coverage: 4.1 in Q20 base.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1016: contig of 1016 bp in length
* 1017 1116: gap of 100 bp
* 1117 2136: contig of 1020 bp in length
* 2137 2236: gap of 100 bp
* 2237 3571: contig of 1335 bp in length
* 3572 3671: gap of 100 bp
* 3672 5272: contig of 1601 bp in length
* 5273 5372: gap of 100 bp
* 5373 6381: contig of 1009 bp in length
* 6382 6481: gap of 100 bp
* 6482 8508: contig of 2027 bp in length
* 8509 8608: gap of 100 bp
* 8609 9886: contig of 1278 bp in length
* 9887 9986: gap of 100 bp
* 9987 11943: contig of 1957 bp in length
* 11944 12043: gap of 100 bp
* 12044 15112: contig of 3069 bp in length
* 15113 15212: gap of 100 bp
* 15213 19323: contig of 4111 bp in length
* 19324 19423: gap of 100 bp
* 19424 24065: contig of 4642 bp in length
* 24066 24165: gap of 100 bp
* 24166 31658: contig of 7493 bp in length
* 31659 31758: gap of 100 bp
* 31759 38737: contig of 6979 bp in length
* 38738 38837: gap of 100 bp
* 38838 46131: contig of 7294 bp in length
* 46132 46231: gap of 100 bp
* 46232 52344: contig of 6113 bp in length
* 52345 52444: gap of 100 bp
* 52445 59763: contig of 7319 bp in length
* 59764 59863: gap of 100 bp
* 59864 66816: contig of 6953 bp in length
* 66817 66916: gap of 100 bp
* 66917 74010: contig of 7094 bp in length
* 74011 74110: gap of 100 bp
* 74111 80599: contig of 6489 bp in length
* 80600 80699: gap of 100 bp
* 80700 89507: contig of 8808 bp in length

* 89508 89607: gap of 100 bp
* 89608 100337: contig of 10730 bp in length
* 100338 100437: gap of 100 bp
* 100438 114298: contig of 13861 bp in length
* 114299 114398: gap of 100 bp
* 114399 128530: contig of 14132 bp in length
* 128531 128630: gap of 100 bp
* 128631 142917: contig of 14287 bp in length
* 142918 143017: gap of 100 bp
* 143018 156243: contig of 13226 bp in length
* 156244 156343: gap of 100 bp
* 156344 172298: contig of 15955 bp in length.

FEATURES

Location/Qualifiers
1..172298
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-57C21"
/clone_lib="RPC1-11 Human Male BAC"
1..1016
/note="assembly_fragment"
1117..2136
/note="assembly_fragment"
2237..3571
/note="assembly_fragment"
3672..5272
/note="assembly_fragment"
5373..6381
/note="assembly_fragment"
6482..8508
/note="assembly_fragment"
8609..9886
/note="assembly_fragment"
9987..11943
/note="assembly_fragment"
12044..15112
/note="assembly_fragment"
15213..19323
/note="assembly_fragment"
19424..24065
/note="assembly_fragment"
24166..31658
/note="assembly_fragment"
31759..38737
/note="assembly_fragment"
38838..46131
/note="assembly_fragment"
46232..52344
/note="assembly_fragment"
52445..59763
/note="assembly_fragment"
59864..66816
/note="assembly_fragment"
66917..74010
/note="assembly_fragment"
clone_end:T7
vector_side:left
74111..80599
/note="assembly_fragment"
clone_end:SP6
vector_side:left
80700..89507
/note="assembly_fragment"
89608..100337
/note="assembly_fragment"
100438..114298
/note="assembly_fragment"
114399..128530
/note="assembly_fragment"
128631..142917
/note="assembly_fragment"
143018..156243
/note="assembly_fragment"
156344..172298

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

Baldwin, J., Barna, N., Beckerly, R., Boguslavskiy, L., Boukhgalter, B., Brown, A., Castle, A., Collangelo, M., Collins, S., Collimore, A., Cooke, P., DeArelano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J., Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 22, 2000 this sequence version replaced gi:7259702.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE JOURNAL COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L1240

Center clone name: 179_B-14

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 168987 bases at least Q40
Consensus quality: 179731 bases at least Q30
Consensus quality: 185225 bases at least Q20
Insert size: 183000; agarose-tp
Insert size: 189119; sum-of-contigs
Quality coverage: 3.9 in Q20 bases; agarose-tp
Quality coverage: 3.8 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 23 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
13 112: contig of 12 bp in length
13 112: gap of 100 bp
113 1767: contig of 1655 bp in length
1768 1867: gap of 100 bp
1868 3764: contig of 1897 bp in length
3765 3864: gap of 100 bp
3865 6037: contig of 2173 bp in length
6038 6137: gap of 100 bp
6138 8137: contig of 2000 bp in length
8138 8237: gap of 100 bp
8238 10419: contig of 2182 bp in length
10420 10519: gap of 100 bp
10520 13978: contig of 3459 bp in length
13979 14078: gap of 100 bp
14079 41398: contig of 27320 bp in length
41399 41498: gap of 100 bp
41499 45869: contig of 4371 bp in length
45870 45969: gap of 100 bp
45970 49945: contig of 3976 bp in length
49946 50045: gap of 100 bp
50046 56380: contig of 6335 bp in length
56381 56480: gap of 100 bp
56481 60955: contig of 4475 bp in length
60956 61055: gap of 100 bp
61056 66082: contig of 5027 bp in length

* 66083 66182: gap of 100 bp
* 66183 74638: contig of 8456 bp in length
* 74639 74738: gap of 100 bp
* 74739 81245: contig of 6507 bp in length
* 81246 81345: gap of 100 bp
* 81346 89538: contig of 8193 bp in length
* 89539 89638: gap of 100 bp
* 89639 102730: contig of 13092 bp in length
* 102731 102830: gap of 100 bp
* 102831 115832: contig of 13002 bp in length
* 115833 115932: gap of 100 bp
* 115933 129142: contig of 13210 bp in length
* 129143 129242: gap of 100 bp
* 129243 146224: contig of 16982 bp in length
* 146225 146324: gap of 100 bp
* 146325 167164: contig of 20840 bp in length
* 167165 167264: gap of 100 bp
* 167265 188461: contig of 21197 bp in length
* 188462 188561: gap of 100 bp
* 188562 191319: contig of 2758 bp in length.

FEATURES source

1. 191319
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3"
/clone="RP11-179B14"
/clone_lib="RPC1-11 Human Male BAC"

misc_feature

1. 12
/note="assembly_fragment"
clone_end:SP6
vector_side:left

misc_feature

113. 1767
/note="assembly_fragment"

misc_feature

1868. 3764
/note="assembly_fragment"

misc_feature

3865. 6037
/note="assembly_fragment"

misc_feature

6138. 8137
/note="assembly_fragment"

misc_feature

8238. 10419
/note="assembly_fragment"

misc_feature

10520. 13978
/note="assembly_fragment"

misc_feature

14079. 41398
/note="assembly_fragment"

misc_feature

41499. 45869
/note="assembly_fragment"

misc_feature

45970. 49945
/note="assembly_fragment"

misc_feature

50046. 56380
/note="assembly_fragment"

misc_feature

56481. 60955
/note="assembly_fragment"

misc_feature

61056. 66082
/note="assembly_fragment"

misc_feature

66183. 74638
/note="assembly_fragment"

misc_feature

74739. 81245
/note="assembly_fragment"

misc_feature

81346. 89538
/note="assembly_fragment"

misc_feature

89639. 102730
/note="assembly_fragment"

misc_feature

102831. 115832
/note="assembly_fragment"

misc_feature

115933. 129142
/note="assembly_fragment"

misc_feature

129243. 146224
/note="assembly_fragment"

misc_feature

146325. 167164
/note="assembly_fragment"

misc_feature

167265. 188461
/note="assembly_fragment"

gene
CDS

/db_xref="taxon:9606"
/chromosome="7"
/map="7q22"
join(1..1531,1635..1820,3728..3843,4264..4328,5271..5419,5858..5923,6280..6442,6962..7135,7228..7319,7475..7527,7697..7738)
/gene="MUC3A"
join(<1..1531,1635..1820,3728..3843,4264..4328,5271..5419,5858..5923,6280..6442,6962..7135,7228..7319,7475..7527,7697..7738)
/gene="MUC3A"
/codon_start=1
/product="intestinal mucin"
/protein_id="BAB12116.1"
/db_xref="GI:9929916"

/translation="MISTVSTSTAITSHFTTSETAVTPTPTVPSLSLTDPTSLR
TLTPSSVGTSLTITDPPSPIDTSLTPRTHLISSPSISQSTETSLGTYSPMT
STVRLRTENTPISSESTSIWIPETPTQPPVLTSATGTQTSPTATVFGSTDS
RSTLHTLTPSTALSTIVTSQVPIIPSTHSSLTQTPPTSPSLTSTSEFTSEFT
RGSTNAILTSFTIWSSTPIIMSSPSSASITPVFTIHSVPSXPFTSENV
GSASITGFLSSSATTSTSSSLTALTETPFYSISLPSLTPCPGTITITVPA
SPTPCVMDPSTPTPLTVPFTTEMTCTPTISITQTLTITVMDTSSMMPES
ESSISPNASSSGTCTVPTNFTVSTRLPTSETLSNSVPLPLPGVSTIPLMKPS
SSLPILRTSSKSTHSPPTRTSETPVATQPTILSRRTITRISQMTTQSTLTIT
AGCTCDNGTWEQGCACLPFGSGRCQLQTRCQNGQWDGLKCCPSTFYGSSCEFA
VEQVDLVETEVEGMEVSDVOQFSPDLNDNTSOAYRDNFKTFWMOQKIFADMQGFTF
KGVEILSRNGSIVVDYLVLLMPFSPQLESEYEQVTKLKEGLQNASQDVNSCODSQ
TLCKPDKSVNNNSKTLTPAACIRRAAPTYEEFYPLVEATRLRCVTKCTSGVDN
ATDCHOGCVLTGCTCRCTSTDPHFSGPRCEVAVHVRALVGLTAGAALLVLLIL
ALGVRAVRSQHWGGRGRSNDOKRWFETHDEEVVGFSGNWGFEDDGTDKDNYFA
LENVDTMKVHKRPMTSSSV"

BASE COUNT 1777 a 2792 c 2126 g 2217 t
ORIGIN

Query Match 5.2%; Score 42.6; DB 85; Length 8912;
Best Local Similarity 46.0%; Pred. No. 0.44; Mismatches 169; Indels 0; Gaps 0;
Matches 144; Conservative 0;

QY 445 acagattggaagtacacaaagtagcattaccacatgagaggagaaacacagagaaa 504
| | | | |
Db 3310 AAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 3251
| | | | |
QY 505 atcttgatgcttcacagacatgcacacacacacacacacacacacacacacacac 564
| | | | |
Db 3250 AAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 3191
| | | | |
QY 565 agccaagctgggagagagataaccacggggcagagggtcaggattctggccctgctgc 624
| | | | |
Db 3190 AGAATTGTGGAGATGAGTGAAAGCAGCGCAGAAAGATAGCGGCTGTGGTCCCT 3131
| | | | |
QY 625 aaactgtgcgttcataac 684
| | | | |
Db 3130 CCTGGCTGCCCATAGTAGGGAGATGTCCATGTCTATATCCCGCAACTGTGACTATGC 3071
| | | | |
QY 685 aatatgtatctctacggttcctctctggggccacacacacacacacacacacacacac 744
| | | | |
Db 3070 TATACTACATGCGCAGGAGGATCAAGGCTGCAGATTGAATTAGATTGCTAATCAGCCCTA 3011
| | | | |
QY 745 atttttaatttt 757
| | | | |
Db 3010 ACTTTTTTTATT 2998

RESULT 15
AF113616/c
LOCUS AF113616 10708 bp DNA PRI 24-NOV-1999
DEFINITION Homo sapiens intestinal mucin 3 (MUC3) gene, partial cds.
ACCESSION AF113616
VERSION AF113616.1 GI:6466800
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 10708)
Crawley,S.C., Gum,J.R. Jr., Hicks,J.W., Pratt,W.S., Aubert,J.P.,
Swallow,D.M. and Kim,Y.S.
Genomic organization and structure of the 3' region of human MUC3:
alternative splicing predicts membrane-bound and soluble forms of
the mucin
Biochem. Biophys. Res. Commun. 263 (3), 728-736 (1999)
99443732
10512748
2 (bases 1 to 10708)
Crawley,S.C., Gum,J.R., Hicks,J.W., Pratt,W.S., Aubert,J.-P.,
Swallow,D.M. and Kim,Y.S.
Direct Submision
Submitted (14-DEC-1998) GI Research, VAMC-SF, 4150 Clement Street,
San Francisco, CA 94121, USA
Location/Qualifiers
1..10708
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q22"
/tissue_type="intestine"
Join(<1..2549,2656..2841,4757..4872,5292..5356,6299..6447,
6886..6951,7307..7469,7989..8162,8255..8346,8502..8554,
8724..9939)
/gene="MUC3"
/product="intestinal mucin 3"
<1..9939
/gene="MUC3"
Join(<1..2549,2656..2841,4757..4872,5292..5356,6299..6447,
6886..6951,7307..7469,7989..8162,8255..8346,8502..8554,
8724..8765)
/gene="MUC3"
/codon_start=2
/product="intestinal mucin 3"
/protein_id="AAF13032.1"
/db_xref="GI:6466801"
/translation="ITITETSHSPSTVTTTITETPSHSPSTVTTTITETPSHS
TPSTSSITTTETSHSPSTSSIRTTETTSYSPSTSSNTITETSHSPSTVITS
ITTTETPSSTPSSTSSITTTETSHSPSTPGTSSITTTETSHSPSTSSITTTET
SHDTPSTSSITTSSTPSSTSLITTTKTTSHSPSTSSITTTETSHSPSTSSITTSARSF
TSSITTTSHNTSSSTTTTSHSTSSSTSSITTTETSHSPSTSSITTTETSHSPSTSSITTT
ETPLHSTPLGTSWTTTATTHGTTHSTTTTSHSTTTTSHSTPGTSSITTTETSEST
PSLSSSTIVSTVSTTATTHGTTHSTTAVTPTVPSLSLTDPTSLTSLRTLPSSVG
TSLTTLTDFPSIPTDLTPRTHLISSPSISQSTETSSILVGTPTMTSTVMTLRL
ITENTPSSSTSIWIPETPTQPPVLTSATGTQTSPTATVFGSTDSSTSLTLTL
TPTALSTIVTSQVPIIPSTHSSLTQTPPTSPSLTSTSEFTSEFTSEFTSEFTSEFT
ILTFSTIWSSTPIIMSSPSSASITPVFTIHSVPSXPFTSTENVSASITGF
PSLSSSATTSTSSSLTALTETPFYSISLPSLTPCPGTITITVPASTDPCE
MDPTEATSPPTPLTVPFTTEMTCTPTISITQTLTITVMDTSSMMPESSTSPNA
SSSGTGCPVNTVFTSRPTSETLSNSVPLPLPGVSTIPLTKMPSSSLPILR
TSSKSTHSPPTRTSETPVATQPTILSRRTITRISQMTTQSTLTITVMDTSSMMPESSTSPNA
TTEWGCACLPFGSGRCQLQTRCQNGQWDGLKCCPSTFYGSSCEFAVFOVDLVY
ETVEGMEVSDVOQFSPDLNDNTSOAYRDNFKTFWMOQKIFADMQGFTFKEVPLSLR
NGSIVVDYLVLLMPFSPQLESEYEQVTKLKEGLQNASQDVNSCODSOTLCFKPDSI
KVNNSKTELTPAACIRRAAPTYEEFYPLVEATRLRCVTKCTSGVDNAIDCHOGC
VLETSQPCRCYSTDTHWFGPRCEVAVHVRALVGLTAGAALLVLLILALGVRAVS
GWGQGRGRSNDOKRWFETHDEEVVGFSGNWGFEDDGTDKDNTFYVALENVDTTMK
VHKRPMTSSSV"
repeat_region 1..1021
/rpt_type=tandem
/rpt_unit=1..51
8053..8127
/gene="MUC3"
/note="Region: membrane spanning domain"
BASE COUNT 2202 a 3424 c 2471 g 2609 t 2 others
ORIGIN

Query Match 5.2%; Score 42.6; DB 89; Length 10708;
Best Local Similarity 46.0%; Pred. No. 0.45;

THIS PAGE BLANK (USPTO,

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 04:11:52 : Search time 6114.61 Seconds
(without alignments)
5537.577 Million cell updates/sec

Title: US-09-402-713A-6
Perfect score: 3582
Sequence: 1 acagaagaatagcaagtgc.....tgattctttgttacaacttt 3582

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues
Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

COMMENT

Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@igr.org

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html
Seq primer: T7
Class: BAC ends.

FEATURES

Location/Qualifiers
Source
1..654
/organism="Homo sapiens"
/db_xref="GDB:754135"
/db_xref="taxon:9606"
/clone="RPCI-11-108L4"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC111 Human Male BAC Library"

BASE COUNT 213 a 100 c 139 g 202 t
ORIGIN

Query Match 17.2%; Score 615.8; DB 227; Length 654;
Best Local Similarity 99.2%; Pred. No. 3.1e-151;
Matches 650; Conservative 0; Mismatches 2; Indels 3; Gaps 3;

QY 1217 tgcacattctcaagacctcaaatgtcattccattcaataataacacagagattacattttttt 1276
DB 654 TCCCATCTTCAAGACCTCAAAATGTCAATTCATTAAATACAGAGATTAATTTTTT 595
QY 1277 ttaacttggaagaattcaatgttaccatgcagctatgggaatttaattacattttttt 1336
DB 594 TTAACCTGGGAAGATCAATGTATACATGCAGCTATGGGAATTAATTACATATTTGTTT 535
QY 1337 tccagtgcgaagactgaacttcctttatccctccctctgtttgtttttttttccagta 1396
DB 534 TCCAGTGCAGAGATGACTAAGTCTTATCCCTCCCTTTGTTGATTTTTCAGTA 475
QY 1397 taaagttaaaatgcttagccttgcctgagctgtatatacag-cacagcctcccccatcc 1455
DB 474 TAAAGTTAAATGCTTAGCTTGTACTAGAGCTGTATACAGCCACAGCCTCTCCCATCC 415
QY 1456 ctccagccttatctgcattcaccaatcaacccctcccataccacaaacaaatcctaa 1515
DB 414 CTCCAGCCTTATCTGTCATCACCATCAACCCCTCCCAT-GCACCTAAACAAATCTAACT 356
QY 1516 tgaattccttgaacatgtcaggacatacattattccttctgcctgagagctcttccct 1575
DB 355 TGTAAATCTCTTGAACATGTGAGG-CATACATATTCTCTCTGCCTGAGAAGCTCTTCTCT 297
QY 1576 gctctttaaatagatgatataagattttgaataagtgtactatcttacttcataa 1635
DB 296 GTCTCTTAATCTAGAATGATGAAGTTTGAATAAGTTGACTATCTTACTTACATGCA 237
QY 1636 aagaaggacacatatgaattcattcatcatcatgacacacacaaactactaaagttaatt 1695
DB 236 AGAAGGACACATATGAGATTTCATCATCATGAGACCAATACTATAAGTGTAAATTT 177
QY 1696 gattataagagtttagataaataatgatgaatgcaagagccacagaggggaattttatgg 1755
DB 176 GATTATAAGAGTTAGATAAATATATGAATCAAGAGCCACAGAGGGAATGTTATGGG 117
QY 1756 gcaacttttagccttggaatgtaggaagcaaaagcaggggaacctcatagtatctataa 1815
DB 116 GCACGTTTGAAGCCTGGGATGTGAAGCAAGGACGAGGAACCTCATAGTATCTTATATA 57

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	615.8	17.2	654	227	AQ319247 RPCI11-10
C 2	536	15.0	546	227	AQ318289 RPCI11-10
C 3	533	14.9	542	227	AQ318297 RPCI11-10
C 4	514.2	14.4	856	21	AI557225 PT2_1_15
C 5	401.2	11.2	441	170	BF858890 RC5-FT019
C 6	394	11.0	402	9	AA578773 nh24a04.s
C 7	362	10.1	394	225	AQ206972 HS_3238.B
C 8	338.4	9.4	346	122	AW948945 QV4-FT000
C 9	331	9.2	451	226	AQ246715 HS_2059.B
C 10	306.2	8.5	332	170	BF858286 RC5-FT019
C 11	260.4	7.3	290	147	BF373619 MR0-FT017
C 12	226.8	6.3	282	147	BF373581 MR0-FT017
C 13	214.4	6.0	226	170	BF858371 RC5-FT019
C 14	167	4.7	167	147	BF373406 IL2-FT015
C 15	160.2	4.5	630	226	AQ302255 HS_3141.B
C 16	86.4	2.4	657	21	AI557495 PT2_1_7.H
C 17	63.2	1.8	1101	219	CNS00039G
C 18	62.6	1.7	600	223	AQ673276 HS_5469.B
C 19	61	1.7	376	223	AQ069980 HS_3027.B
C 20	59.6	1.7	1067	219	CNS000C08
C 21	58.8	1.6	997	219	CNS000TE
C 22	56.8	1.6	1001	219	CNS01400
C 23	56.2	1.6	1101	219	CNS000D1
C 24	55	1.5	463	257	B81447 CIT-HSP-205
C 25	52.8	1.5	1101	219	CNS0100X
C 26	52.6	1.5	1101	219	CNS0182P
C 27	51.4	1.4	987	219	CNS00418
C 28	51	1.4	1034	220	CNS02N17
C 29	50.2	1.4	1101	219	CNS00039E
C 30	49.6	1.4	958	221	CNS03CMC
C 31	49.2	1.4	1101	219	CNS0105X
C 32	49	1.4	1101	219	CNS017KX
C 33	49	1.4	1204	219	CNS016E2
C 34	48.6	1.4	1090	221	CNS04U88
C 35	48.4	1.4	1086	192	AQ020508 Mus muscu
C 36	47.6	1.3	529	225	AQ173559 HS_3202.A
C 37	47.4	1.3	1458	192	AQ018211 Mus muscu
C 38	47.2	1.3	586	235	AQ928836 RPCI-23-2
C 39	47.2	1.3	659	242	AZ358277 1M0100F09
C 40	47.2	1.3	677	242	AZ357710 1M0099C05
C 41	46.8	1.3	654	223	AQ046642 RPCI11-35
C 42	46.8	1.3	713	240	AZ253521 RPCI-23-4
C 43	46.8	1.3	854	140	BE777727 601348546
C 44	46.8	1.3	1101	219	CNS00067
C 45	46.6	1.3	421	163	BE137644 ug66f11.y

ALIGNMENTS

RESULT 1
ACQ319247/c
LOCUS AQ319247 654 bp DNA GSS 06-MAY-1999
DEFINITION RPCI11-108L4.TV RPCI-11 Homo sapiens genomic clone RPCI-11-108L4,
DNA sequence.
ACCESSION AQ319247
VERSION AQ319247.1 GI:4052212
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 654)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
TITLE Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL Unpublished (1998)

```

QY 1816 tatacttcattctctatctctatcaaatatccaaagctttttcacagaattc 1870
|||||
Db 56 TATACTTCAATTTCTATCTCTATCAAAATATCAACAGCTTTTCAGAGAAATC 2

RESULT 2
AQ318289/c
LOCUS AQ318289 546 bp DNA GSS 04-MAY-1999
DEFINITION RPCI11-108A18.TV RPCI-11 Homo sapiens genomic clone RPCI-11-108A18,
DNA sequence.
ACCESSION AQ318289
VERSION AQ318289.1 GI:4050259
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 546)
Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
Use of human BAC End Sequences for Sequence-Ready Map Building
Unpublished (1998)
Other_GSSs: RPCI11-108A18.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html
Seq primer: T7
Class: BAC ends.
FEATURES
Location/Qualifiers
1..546
/organism="Homo sapiens"
/db_xref="GB:7541105"
/db_xref="taxon:9606"
/clone="RPCI-11-108A18"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
RPCI11 Human Male BAC Library"
BASE COUNT 150 a 106 c 116 g 172 t 2 others
ORIGIN

Query Match 15.0%; Score 536; DB 227; Length 546;
Best Local Similarity 98.7%; Pred. No. 3.2e-130;
Matches 539; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2287 aaagcagctggaatggacacacaaatgcataatcttaactctaccatcagttaca 2346
|||||
Db 546 AAAGCAGCTGGAATGGACAAACCAATATGCATAATCTCAACNTCTACCATCAGCTACA 487

QY 2347 cactgttgacatatattgttagaagcacctctgcatttgggtttctttaagcaaaata 2406
|||||
Db 486 CACTGCTTGACATATATTGTTAGAACACCTCGCAITTTGGGGTCTCTTAAGCAAAATA 427

QY 2407 cttgcattaggctcagctg99g9gctgcatcagcg9g9gttgagaaatattcaattctca 2466
|||||
Db 426 CTTGCAATTAGGCTCAGCTGGGGCTGTGCATCAGNCGGTTTGAGAAATATTCAATTTCTCA 367

QY 2467 gcagaagccagaatttgattccctcatcttttagaatacatttaccaggtttggagagg 2526
|||||
Db 366 GCAGAAGCCAGAAATTTGAATTCCTCATCTTTTAGGAATCATTTACCAGGTTTGGAGAG 307

```

```

QY 2527 attcagacagctcaggtgttttcaactaagtctctgaacttctgtccctcttggttca 2586
|||||
Db 306 ATTCAAGACAGCTCAGGTGCTTTTCACATAATGCTCTGAACATTCCTGCTCCCTTTGGTTCA 247

QY 2587 tggatagtcocataataatgtttattcttttgaactgatgctcatagagagaataaagaa 2646
|||||
Db 246 TGGATAGTCCAATATAATATGTTATCTTTGAAGTATGCTCATAGGAGAGATAAAGAA 187

QY 2647 ctctgagtatatacaattaggattcaagaataattagatttaagctcacacgtgtc 2706
|||||
Db 186 CTCTGAGTGATATCAACATTAGGGATTCAAGAAATATTAGATTAAAGCTCACACGTGGTC 127

QY 2707 aaaggaaccaagatacacaagaactctgagctgctatcgctcccatctctgtgagccaca 2766
|||||
Db 126 AAAAGGAACCAAGATACAAAGAACTCTGAGCTGCTATGCTCCCATCTCTGTGAGCCACA 67

QY 2767 accaacagcaggagcccaacgcatgtctgagatccttaaatcaaggaaacacgtgtcatga 2826
|||||
Db 66 ACCAACAGCAGGAGCCCAACGATGCTGTGAGATCTTTAAATCAAGGAAACCAAGTGTCTATGA 7

QY 2827 gttgaa 2832
|||||
Db 6 GTTGAA 1

RESULT 3
AQ318297/c
LOCUS AQ318297 542 bp DNA GSS 04-MAY-1999
DEFINITION RPCI11-108A20.TV RPCI-11 Homo sapiens genomic clone RPCI-11-108A20,
DNA sequence.
ACCESSION AQ318297
VERSION AQ318297.1 GI:4050267
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 542)
Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
Use of human BAC End Sequences for Sequence-Ready Map Building
Unpublished (1998)
Other_GSSs: RPCI11-108A20.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html
Seq primer: T7
Class: BAC ends.
FEATURES
Location/Qualifiers
1..542
/organism="Homo sapiens"
/db_xref="GB:7541107"
/db_xref="taxon:9606"
/clone="RPCI-11-108A20"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
RPCI11 Human Male BAC Library"
BASE COUNT 151 a 106 c 118 g 167 t
ORIGIN

```

Query Match 14.9%; Score 533; DB 227; Length 542;
Best Local Similarity 100.0%; Pred. No. 2e-129;
Matches 533; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2303 gacacacacataatgataaaatctaaactcctaccatcagctacacacacgcttgacatata 2362
DB 542 GACAAACCAATATGATCAAAATCTAACTCCTACCATCAGTACACACATGCTTGACATATA 483
QY 2363 ttgttagagacacctgcgatttgggttctcttaagcaaaatcttcattgattgctcca 2422
DB 482 TTGTTAGAGACACCTCGCATTTGCTGGTTCCTTAAAGCAAAATATCTTGCATTAGTCTCA 423
QY 2423 gctgggggtgtgcatcagcggtttgagaaatattcattctcagcagaagcagaattt 2482
DB 422 GCTGGGGCTGTGCATCAGCGGTTTGAAGAAATATTCAATTCACAGAAAGCCAGAAATTT 363
QY 2483 gaattccctcatcttttaggaataattaccaggtttgagaggattcagacagctcagg 2542
DB 362 GAATTCCTCTATCTTTTAAAGGAATCAATTTACACAGGTTTGGAGAGGATTGAGACAGCTCAGG 303
QY 2543 tgccttactaatgtctctgaactctgtccctcttctgttctcagtgatagcctaataa 2602
DB 302 TGCCTTCTACTAATGTCTCTGAACCTTCTGTCCCTCTTGTGTTTCATGGATAGTCCCAATAA 243
QY 2603 taattgtatctttgaactgactgctcatcaggagagagataaagaactctgagtgatataca 2662
DB 242 TAATGTTATCTTTGAACCTGATCTCATAGGAGAGAGATAATAAGAACTCTGAGTGATATCAA 183
QY 2663 cattaggattcaagaagaattattagatttaagctcacactgctgcaaaaggaaaccaagata 2722
DB 182 CATTTAGGAGATCAAAAGAAATATTAGATTAAAGCTCACACTGGTCAAAAGGAACCAAGATA 123
QY 2723 caaagaactctgagctgctcatcgtcccatctctgtgagccacaacacacagcagacc 2782
DB 122 CAAGAAGCTGTGAGCTGTGATCGTCCCATCTCTGTGAGCCCAACCAACAGCAGACCC 63
QY 2783 aacgcatgctgagatccttaaatcaagaaacacagtgctcatgagttgaattc 2835
DB 62 AACGCATGCTGAGATCTCTTAATCAAGGAACCAAGTGTGATGAGTGAATTC 10

RESULT 4
AI557225 856 bp mRNA EST 09-AUG-1999
LOCUS PT2.L15_B05.r tumor2 Homo sapiens cDNA 3', mRNA sequence.
DEFINITION AI557225
ACCESSION AI557225
VERSION AI557225.1 GI:4489588
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 856)
AUTHORS Huang,G.M., Ng,W.L., Parks,J., He,L., Liang,H.A., Gordon,D., Yu,J.
and Hood,L.
TITLE Prostate cancer expression profiling by cDNA sequencing analysis
JOURNAL Genomics 59 (2), 178-186 (1999)
MEDLINE 99339982
COMMENT Contact: Guyang Matthew Huang
Leroy Hood
University of Washington
Department of Molecular Biotechnology, Box 357730, University of
Washington, Seattle, WA 98195
Tel: 5106280100
Fax: 5106280108
Email: huanggm@yahoo.com.
Location/Qualifiers
1 . 856
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="tumor2"
/note="Organ: Prostate; Vector: pBluescript; Directional

cDNA library was constructed using Lambda ZP II kit (Stratagene). mRNA was extracted from a frozen prostate tumor tissue (Mayo Clinics)."

BASE COUNT 237 a 191 c 180 g 214 t 34 others
ORIGIN

Query Match 14.4%; Score 514.2; DB 21; Length 856;
Best Local Similarity 97.1%; Pred. No. 2.1e-124;
Matches 575; Conservative 0; Mismatches 11; Indels 6; Gaps 5;

QY 1352 actaagtcctttatccctcccttctgttggattttttccagtgataaaagttaaaatgct 1411
DB 18 ACCTAGCCCTTTAATCCCTCCCTCTTGTGATTTTTCAGTATAAAAGTTAAAAATGCT 77
QY 1412 tagccttgactgagcgtgtatcacag-cacagcctctcccccctccctccagcctatctg 1470
DB 78 TAGCCTTGTACTGAGCGTGTATACAGCCACAGCCCTCTCCCATCCCTCCAGCCTTATCTG 137
QY 1471 tcatcaccatcaacccctccctccatcaccacccaacaaaatctaaacttgtaattccttgaac 1530
DB 138 TCATCACCATCAACCCCTCCCAT-GCACCTAAACAAAATCTAACTTGTAATTCCTTTGAAC 196
QY 1531 atgtcaggacatacatattctcttctgctgagagctcttctctgtctctcttaaatctag 1590
DB 197 ATGTCAGG-CATACATTAATTCCTTCTGCTGAGAAAGCTCTCTCTGCTCTTTAAATCTAG 255
QY 1591 aatgatgtaaagtttgaataaagttagctatcttcttctcatcgaagaagggacacatat 1650
DB 256 AATGATGTAAAGTTTGAATAAGTTGACTATCTTACTTCATGCAAGAAGGGACACATAT 315
QY 1651 gagattcatcatcagatgagacagcaaaactaataaagtgttaatttgattataagagttta 1710
DB 316 GAGATTCAATCATCATGAGACAGCAAAATCTAAAGTGTAAATTTGATTATTAAGAGTTTA 375
QY 1711 gataatatatgaatgcaagagccacagaggggaatgtttatggyggaagctttgtaagcc 1770
DB 376 GATAAATATATGAAATGCAAGAGCCACAGAGGAATGTTTATGGGACGCTTTGTAAGCC 435
QY 1771 tgggattgaaagcaagcaggaacccctcatgattcttataataataataataataataata 1830
DB 436 TGGGATGTGAAGCAAGGCGAGGACCTCATAGTATCTATATAATATATCTCATTTCTC 495
QY 1831 tatctctatcaaatatccaaagcttttccagaatctcatgagtcagtcgcaatcccca 1890
DB 496 NATCTCTATCAATATATCAACAAAGCTTTTCACAGAAATTCATGCGAGTCAATCCCAAA 555
QY 1891 ggtaacctttat-ccatttcattggtgagtcgcttt--agaattttggcaaa 1939
DB 556 GGGAACTTTATCCCATTTTCATGCTGAGTGGCGCTTNAAGAAATTTTGGNAAA 607

RESULT 5
BF58890/c 441 bp mRNA EST 16-JAN-2001
LOCUS RCS-FT0194-071200-023-G11 FT0194 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF58890
ACCESSION BF58890
VERSION BF58890.1 GI:12246634
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 441)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE
COMMENT

20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC5&t2=RC5-PT0194-
071200-023-G1&t3=2000-12-07&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 3
High quality sequence stop: 416.

FEATURES
source

1. .441
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="PT0194"
/dev_stage="Adult"
/note="Organ: prostate_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
142 a 67 c 97 g 135 t

BASE COUNT
ORIGIN

Query Match 11.2%; Score 401.2; DB 170; Length 441;
Best Local Similarity 98.6%; Pred. No. 1.1e-94; Mismatches 3; Indels 3; Gaps 3;
Matches 436; Conservative 0; Mismatches 3; Indels 3; Gaps 3;
QY 1315 aattatcatatttggtttccagtgcaagatgactaaagtcttattccctccct 1374
|||||
Db 441 AATTAAATACATATTGTTTCCAGTGCAAGATGACTAAGTCTTATCCCTCCCT 382
QY 1375 ttattgattttttccagtaaaagttaaatgottagccttgtagcgtgata 1434
|||||
Db 381 TTGTTTGAATTTTTTCCAGTATAAAGTTAAATGCTTAGCCTTGACTGAGGCTGTATA 322
QY 1435 cag-cacagcctctccccctccctccagccttatctgtcatcacccatccat 1493
|||||
Db 321 CAGCCACAGCCTCTCCCATCCCTCCAGCCTTATCTGTATCACCATCAACCCCTCCCAT 262
QY 1494 accacctaataaaatcctaacttgtaattccttggaacatgctcaggacacattattcct 1553
|||||
Db 261 -GCACCTAAACAAATCTAATCTTGAATCTTGAACATGTCAGG-CATACATTATTCCT 204
QY 1554 tctgctgagaagctctctcttctcttaaatctagaatgtagtaaaagtttgataag 1613
Db 203 TCTGCTGGAAGCCTCTCTCTCTCTTAAATCTAGAATGATGTAAGTTTGAATAAG 144
QY 1614 ttgactatcttactcatgcaagaaggacacacatgatgattcatcatcacatgagaca 1673
|||||
Db 143 TTGACTATCTTACTTCATGCAAGAAGGACACATATGAGATTTCATCATCATGAGACA 84
QY 1674 gcaataactaaagttaatttgattataagaggtttagataataatatataatgcaagag 1733
|||||
Db 83 GCAATATCCCAAGTCTAATTTGATTATAAGAGCTTTAGATAAATATAGAAATGCAAGAG 24
QY 1734 ccacagaggaagtgtttatggg 1755
|||||
Db 23 CCACAGAGGAATGTTTATGGG 2

RESULT
AA578773
LOCUS

6
AA578773 402 bp mRNA EST 12-SEP-1997

DEFINITION

nh24a04.s1 NCI_CGAP_Pr1 Homo sapiens cDNA clone IMAGE:953262, mRNA
sequence.
ACCESSION AA578773
VERSION AA578773.1 GI:2356957
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 402)
REFERENCE
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chusqui, M.D.
, Michael Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
DNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/Db/rp/image/image.html
Insert Length: 565 Std Error: 0.00
Seq primer: -40m13 fwd. Et from Amersham
High quality sequence stop: 373.

FEATURES
source

1. .402
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:953262"
/clone_lib="NCI_CGAP_Pr1"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/note="Vector: PAMP10; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was primed with oligo(dt)17 on 50 ng of
DNase-treated, total cellular RNA obtained from 5,000-10
,000 microdissected, histologically normal prostate
epithelial cells. Double-stranded cDNA was ligated to
EcoRI adaptors, 5 cycles of PCR applied to the cDNA with
an adaptor-specific primer, and the resulting PCR product
subcloned into PAMP10 by the UDG-cloning method (Life
Technologies). Average insert size is 600 bp. NOTE: Not
directionally cloned. This library was constructed by
David Krizman."
128 a 86 c 84 g 104 t

BASE COUNT
ORIGIN

Query Match 11.0%; Score 394; DB 9; Length 402;
Best Local Similarity 98.8%; Pred. No. 8.4e-93;
Matches 397; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 582 ggaatggcaggaacacagatccctgtgtgattatttatttgaacgggattcacagattga 641
|||||
Db 1 GGTCCAGCAGGAAACAGATCCTGTGTGATATATATTGACGGGATTACAGATTGA 60
QY 642 aatgaagtccaaaagtgcattaccattgaggaagaaacagacagagaaatcttgatgg 701
|||||
Db 61 AATGAAGTCACAAAGTGAGCATTTACCAATGAGAGAAACAGACAGAGAAATCTTGATGG 120
QY 702 cttcacaagacatgcaacaaataactgtatgatcattgacatgagcagcagactg 761
|||||
Db 121 CTTCAACAGCATGCAACAAACAAATGGAATCTGTATGATGATGAGGAGCCAGACTG 180
QY 762 gggagagataaccacggggcagaggggtcaggattcttgcctgtgcctaaactgtgcg 821
|||||
Db 181 GGGAGGAGATAACACACGGGCGCAGGGTCAGGATTTGGCCCTGCTGCCTAAACTGTGCG 240
QY 822 ttcataacaaatcatttcattatatttctaaccctcaaaacaaagtgttgaatactgat 881
|||||

Db 241 TTCATACCAAAATCATTTTCATATATTTCTTAACCCCTCAAAACAAAGCTGTGTCTAATATCTGAT 300

Qy 882 ctctacgattctcttgccgcaacattctccatatatccagccacacactctttttaata 941

Db 301 CTCACGGTTCCTCTGGGCGCAACATTTCTCATATATCCAGCCACACTCATTTTAAATA 360

Qy 942 tttagtctccagatctgtactgtgacctttctacactgtaga 983

Db 361 TTTAGTCTCCAGATCTGTACTGTGACCTTCTACACTGTAGA 402

RESULT 7

LOCUS AQ206972/c 394 bp DNA GSS 17-SEP-1998

DEFINITION HS_3238_B1_G11_MR_CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3238 Col=21 Row=N, DNA sequence.

ACCESSION AQ206972

VERSION AQ206972.1 GI:3617542

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 394)

AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE 99380589

COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3238 row: N column: 21
Class: BAC ends
High quality sequence stop: 394.
Location/Qualifiers
1. .394
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Plate=3238 Col=21 Row=N"
/sex="male"
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 134 a 62 c 93 g 103 t 2 others

ORIGIN

Query Match 10.1%; Score 362; DB 225; Length 394;
Best Local Similarity 97.7%; Pred. No. 2,3e-84;
Matches 387; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

Qy 892 ctctctgggcccacattctccatatatccagccacactctttttaataatttagtccc 951

Db 394 CCGTCTGGGCNACATTTCTCCATATATCCAGCCACACTCATTTTAAATATTAGTCCC 335

Qy 952 agatctgtactgtacctttctacactgtagataacattactctattttgttccaagacc 1011

Db 334 AGATCTGTACTGTGACCTTCTACACTGTGAGTAATAACATTTACTATTGTTCGAAGACC 275

Qy 1012 ctctctgttgcctctaataatagctactgtttttctaaaggagtgcttgcgccagg 1071

Db 274 CTCGTGTGTGTCCTTAATATAGTACTGTCTTTTCTAAGGAGTGTCTGGCCAGG 215

Qy 1072 ggaatctgaaacaggctgggaagcatctcaagatctcttccagggttatacttactagcac 1131

Db 214 GGATCTGTGAACAGGCTGGGAAGCATCTCAAGATCTTTCCAGGTTATATCTACTAGCAC 155

Qy 1132 acagcatatcatcagcagtggaattatctaatcaacatcatcctcagtgctcttgccca 1191

Db 154 ACAGCATGATCATTTACGGAGTGAAATTATATATCAACATCATCTCTCTTTTGCCCA 95

Qy 1192 lactgaaattcatctccacacttttgcacatttccacacacctcaagacacctcaaaatgtctaccatt 1251

Db 94 TACTGAATTCATTTCCACATTTTCTGTG-CCATTCCTCAAGA-CTCAAAATGTCTATTCATT 37

Qy 1252 aatcacagaggaataacttttttttttaacctggaa 1287

Db 36 AATATCACAGGATTAACCTTTTTTTTTTATACCTGCAA 1

RESULT 8

LOCUS AW948945 346 bp mRNA EST 31-MAY-2000

DEFINITION QV4-FT0004-130500-212-d11 FT0004 Homo sapiens cDNA, mRNA sequence.

ACCESSION AW948945

VERSION AW948945.1 GI:8126719

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 346)

AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-QV4-FT0004-130500-212-d11&t3=2000-05-13&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 344.
Location/Qualifiers
1. .346
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="FT0004"
/dev_stage="Adult"
/note="Organ: prostate_tumor; Vector: puc18; Site_1: Sma1; Site_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 104 a 81 c 63 g 98 t

ORIGIN

Query Match 9.4%; Score 338.4; DB 122; Length 346;
Best Local Similarity 99.7%; Pred. No. 3.7e-78;
Matches 339; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2591 tagtccaataaataatgttattctttgaactgatgcctcagagagaataataagaactct 2650

```

Db      7 TAGTCCAAATAAATGTTACCTTTGAACTGATGCTCATAGGAGAAATATAGAACTCT 66
QY      2651 gagtgtatcaacattaggagattcaagaataattagatttaagctcacactggtcaaaa 2710
Db      67 GAGTGATATCAACATTAGGAGATTCAAGAAATATTAGATTAAAGCTCACACTGGTCAAAA 126
QY      2711 ggaaccaagatacaagaagactctgagctgtcatctgtcccatctctgtgagccacaacca 2770
Db      127 GGAACCAAGATACAAAGAACTCTGAGCTGTGTCATCGTCCCATCTCTGTGAGCCACAACCA 186
QY      2771 acagcaggaccacacacatctgaaatccttaaatcaagaaacacagctgtcatgattg 2830
Db      187 ACAGCAGGACCAACGATCTGTAGATCTCTTAATCAAGAAACAGTGTGATGAGTTG 246
QY      2831 aattctctattatgtagctagctctcttgccatctctctgctctctctcttgacacatt 2890
Db      247 AATTCTCTATTATGATGCTAGCTCTTGCCATCTCTGGCTCTCTCTGTGACACATATT 306
QY      2891 agctttagcttctgttcacagactttattctttctcc 2930
Db      307 AGCTTAGGCTTTGCTTCCAGCACTTTTATCTTTTCTCC 346

RESULT 9
AQ246715 451 bp DNA GSS 06-OCT-1998
LOCUS HS_2059_B2_E08_T7 CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate=2059 Col=16 Row=J, DNA sequence.
ACCESSION AQ246715
VERSION AQ246715
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 451)
Mahairs,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairs GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2059 row: J column: 16
Class: BAC ends
High quality sequence stop: 451.
LOCATION/Qualifiers
1. .451
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="plate:2059 Col=16 Row=J"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in,
E-Coli DH10B"
BASE COUNT 133 a 101 c 86 g 130 t 1 others
ORIGIN

Query Match 9.2%; Score 331; DB 226; Length 451;
Best Local Similarity 90.1%; Pred. No. 3.6e-76;
Matches 411; Conservative 0; Mismatches 36; Indels 9; Gaps 5;
QY 1908 tcattggtgagtcgcttttagaattttggcaaatcatactagctggtcacttatctcaactttga 1967

```

```

Db      2 TCATGTGAGTGGCGCTTTAGAATGTTGGCAAATCATACTGCTCACTTATCTCAACTTTGA 61
QY      1968 gatgtgttctctcttagttaattgaagaataaggcaactcttgtgagccacttttag 2027
Db      62 GATGTGTGTGCGCTTTAGTGAATTTGACAGAAATACGGCACTCTTTGAGCCACTTTTAGG 121
QY      2028 gttcactcctggcaataaagaattttacaaagagctactcaggcaggtgtttaagagctc 2087
Db      122 GCTCACTCCTCGCAATAAAGAATATACANAGAGCTACTCAGACCAGCTGTAGAGCTCT 181
QY      2088 tgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2146
Db      182 GGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 236
QY      2147 ccattatttccagactc-aaacaagcatgttttcaaatggcactatgagctgccaatgat 2205
Db      237 CCATTATTCTAGCTTTAAACCAAGCATGTTTCAATGGCCTATGAGCTGCGCAATGAT 236
QY      2206 gtatcaccccatatctcattattctccagtaaatgtgataaatgtcatctgtttaaca 2265
Db      297 GTATCACCAACCATATCTCATTTATCTCCAGTATATGTGATAAATAATGTATCTATG- 355
QY      2266 taaaaaagtgtgacttcacaaagcagctggaaatggcaacacacaaatgataaatac 2325
Db      356 TATACAAAGTTCGAC-TCACAGAAAGCAGCTGTGAATGCACACCAACCAATATGCATATC 414
QY      2326 taactctaccatcagctcacactgcttgacatat 2361
Db      415 TAATCTCTACCATCAGCTACACACTGCTTGACATAT 450

RESULT 10
BF858286 332 bp mRNA EST 16-JAN-2001
LOCUS RCS-FT0193-201100-012-D06 FT0193 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF858286
ACCESSION BF858286
VERSION BF858286.1 GI:12246030
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 332)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?ti=RC5&t2=RC5-FT0193-
201100-012-D06&t3=2000-11-20&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 332.
LOCATION/Qualifiers
1. .332
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="FT0193"
FEATURES
source

```

/dev_stage="Adult"
/note="Organ: prostate_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 90 a 84 c 64 g 94 t
ORIGIN

Query Match 8.5%; Score 306.2; DB 170; Length 332;
Best Local Similarity 99.0%; Pred. No. 1.1e-69;
Matches 308; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 714 tcacaaacaaatgaataactgtgatgacatgaggaagcgaagctggggagagataa 773
DB 12 TGAACAACAAATGGAATCTGTGATGACATGAGGCGCAAGCTGCGGAGAGATAA 71
QY 774 ccacggggcagagggctcaggattctggccctgctgctaaactgtgcgttcataacccaa 833
DB 72 CCACGGGGCAGAGGGTCAGAGATCTGGCCCTGCTGCTAAACTGCGTTTACACCAA 131
QY 834 tcatttcatttctaaccctcaaaacaaagctgtgttaataatctgatctctacggttcc 893
DB 132 TCATTTTCATATTCTAACCTCAAAACAAAGCTGTTGTAATATCTGATCTACGGTTCC 191
QY 894 ttctggggcccaattctccatatatccagccacacactcatttttaattatgttccgg 953
DB 192 TTTCTGGGCCCAATCTCCATATATCCAGCCACACTCATTTTATATTTAGTTCCAG 251
QY 954 atctgtactgtacattcttcacgttagaataaactactcattttgttcaaaacccct 1013
DB 252 ACTGTACTGTGACCTTCTACGCTGTAGATAAACATTACTATTTGTTTCAAGACCT 311
QY 1014 tcgtgtgtgtg 1024
DB 312 TCGTGTTCAG 322

RESULT 11
BF373619 290 bp mRNA EST 24-NOV-2000
LOCUS MRO-FT0175-310800-106-h09 FT0175 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF373619
ACCESSION BF373619
VERSION BF373619.1 GI:11335644
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 290)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MRO&t2=MRO-FT0175-
310800-106-h09&t3=2000-08-31&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 290.
Location/Qualifiers

FEATURES

source
1..290
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="FT0175"
/dev_stage="Adult"
/note="Organ: prostate_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 78 a 73 c 53 g 86 t
ORIGIN

Query Match 7.3%; Score 260.4; DB 147; Length 290;
Best Local Similarity 97.8%; Pred. No. 1.3e-57;
Matches 264; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 762 gggaggagataaccacggggcagaggtcaggattctggccctgctgctaaactgtgcg 821
DB 21 GGGAGGAGATACCAGGGGGCAGAGGGTCAGGATTCGTGCCCTGTGCCCTAACTGTGGC 80
QY 822 ttataaccacaaatcttcatatttctaaccctcaaaacaaagctgtgttaataatctgat 881
DB 81 TTCATAACCAATCATTTTCATATTTCATACCCCTCAAAACAAAGCTGTTGTAATATCTGAT 140
QY 882 ctctacgggttccttctggggcccaacattctccatatatccagccacacacttttaata 941
DB 141 CTCTAGGTTCTCTTGGGCCCAACATTCCTCATATATCCAGCCACACCATTTTAAATA 200
QY 942 tttagttccagatctgtactgtgaccttctcacctgttagaataacattactcattttg 1001
DB 201 TTTAGTTCACAGATCTGTACTGTGACCTTCTACACTGTAGATAACATTACTCATTTGG 260
QY 1002 ttcaagacccttcgtgtgtgcgtcctaata 1031
DB 261 TTTCAAGACCCCTCGTGTGGCTAAGCACTA 290

RESULT 12

BF373581/c 282 bp mRNA EST 24-NOV-2000
LOCUS MRO-FT0175-210800-101-d05 FT0175 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF373581
ACCESSION BF373581
VERSION BF373581.1 GI:11335606
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 282)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil

Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR06t2-MR0-F70175-210860-101-d056t3-2000-08-21&t4=1>)
Seq primer: puc 18 forward
High quality sequence start: 35
High quality sequence stop: 282.

FEATURES

```

FEATURES
source
      .282
Location/Qualifiers
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone_lib="FT0175"
  /dev_stage="Adult"
  /note="Organ: prostate.tumor; Vector: puc18; Site_1: Sma1;
  Site_2: Sma1; A mini-library was made by cloning products
  derived from ONESTES PCR (U.S. Letters Patent application
  No. 196,716 - Ludwig Institute for Cancer Research)
  profiles into the pUC 18 vector. Reverse transcription of
  tissue mRNA and cDNA amplification were performed under
  low stringency conditions."
      85 a 51 c 70 q 76 t
BASE COUNT

```

BASE COUNT	85 a	51 c	70 g	76 t
ORIGIN				
Query Match		6.3%	Score 226.8;	DB 147; Length 282;
Best Local Similarity		91.3%;	Pred. No. 9.5e-49;	
Matches 252; Conservative		0;	Mismatches 22;	Indels 2; Gaps 1;

QY	755	caagctggggagagataaacaccaggggcagaggggtcagattctggccctgtgcctaaa	814
Db	282	CAAGCTGGGAGAGAGATAAACACGGGGCAGAGGGTCAGGATTCGGCCCTGTCCTAAA	223
QY	815	ctgtcggtcatcaaccaaatcattcatattctaaccctcaaaacaaagctgtgtaat	874
Db	222	CTGTGGGTTCAATAACCAATCATTTCTAACCCCTCAAAACAAGCTGTGTGAAT	163
QY	875	atctgatctctacaggttctctctgtggcccaacattctccatatatccagccacactcatt	934
Db	162	ATCTGATCTCTACGGTTCCTCTGGGGCCCAACATTCCTCATATATCCAGCCACACTCAT	103
QY	935	ttaaattatttagttcccaatctgactgtgaccttc--tacactgtagaataacattta	992
Db	102	TTTAAATATTTAGTTCAGATCTGTACTGTGACCTTCTCTACATCTGTAGAATAACATTA	43
QY	993	ctcatattgttcaaaagacctctgtgtgtgccta	1028
Db	42	CTCATTTGTTCAAAAGACCTTCACCTTGTCTAGCGACTA	7

RESULT	13
BF858371/C	
LOCUS	226 bp mRNA EST 16-JAN-2001
DEFINITION	RC5-FT0193-211100-012-b11 FT0193 Homo sapiens cDNA, mRNA sequence.
ACCESSION	BF858371
VERSION	BF858371.1 GI:12246115
KEYWORDS	EST.
SOURCE	human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 226)

REFERENCE

AUTHORS

Dias Neto, E., Garcia Correa, R., Verjovsky-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE	
COMMENT	

20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC3&t2=RC5-FW0193-211100-012-E11&t3=2000-11-21&t4=1>)
Seq primer: puc 18 forward
High quality sequence stop: 226.

FEATURES

FEATURES	nryn quality sequence stop. 220.		
source	Location/Qualifiers		
	1.	.226	
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone_lib="FT0193"		
	/dev_stage="Adult"		
	/note="Organ: prostate_tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."		
BASE COUNT	79	a	29 c
ORIGIN		59	g
		59	t

	Query Match	6.0%;	Score 214.4;	DB 170;	Length 226;
	Best Local Similarity 99.5%;	Pred. No. 1.7e-45;			
	Matches 215; Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;	
Qy	809	cctaaactgctgcgttcataaccaaatcatttctatttctaaccctcaaaaagaagctgt	868		
Db	226	CCTAAACTGTCGTTCATACCAAAATCATTTTCATTTCCTCNAACAAGCTGT	167		
Qy	869	tgtaatatctgatctctacgggttccttcttgggcccaaacattctccatatatccaggccaaca	928		
Db	166	TGTAATAATCTGATCTACGGTTCCTCTTGCGGCCAACATTTCTCATATATCCAGCCACA	107		
Qy	929	ctcaatttttaatatattagtctccagactctgactgtgaaccttctcacacttgaagaataac	988		
Db	106	CTCAFTTTTTATATTTAGTTCCCAGATCTGTACHTGTGACCCTTTCTACACTGTAGAATAAC	47		
Qy	989	attaactcatttgttccaagaagccccttcogtggctgtg	1024		
Db	46	ATTACTCATTTTGTTCAAAGACCCCTTCGTGTGCAG	11		

RESULT	14
BF373406	
LOCUS	BF373406 167 bp mRNA
DEFINITION	IL2-FT0159-070800-120-H01 FRO159 Homo sapiens EST
ACCESSION	BF373406
VERSION	BF373406.1 GI:11335431
KEYWORDS	EST.
SOURCE	human.
	24-NOV-2000 cdna, mRNA sequence.

SEQUENCE TAGS	ORGANISM	REFERENCE AUTHORS	TITLE
Homo sapiens			Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Primates; Catarrhini; Homnidae; Homo.			
1 (bases 1 to 167)			
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R., Nagai, M. A., da Silva W. Jr., Zago, M. A., Bordin, S., Costa, F. F., Goldman, G. H., Carvalho, A. F., Matsukuma, A., Baia, G. S., Simpson, D. H., Brustein, A., deOliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and Simpson, A. J.			

THIS PAGE BLANK (USP)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 07:25:42 ; Search time 176.8 Seconds
(without alignments)
3835.478 Million cell updates/sec

Title: US-09-402-713A-6
Perfect score: 3582
Sequence: 1 acgaagaataagcaagtgc.....tgattctttgttacaacttt 3582

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	77.4	2.2	7218	1 US-08-232-463-14	Sequence 14, Appl
c 2	57.4	1.6	7218	1 US-08-232-463-14	Sequence 14, Appl
c 3	43	1.2	10968	4 US-08-680-327-2	Sequence 2, Appl
c 4	43	1.2	10968	4 US-09-228-246-1	Sequence 1, Appl
c 5	41.6	1.2	10409	3 US-08-772-440-33	Sequence 33, Appl
c 6	41.4	1.2	5475	2 US-08-680-327-1	Sequence 1, Appl
c 7	41.4	1.2	5475	4 US-09-228-246-3	Sequence 3, Appl
c 8	40.6	1.1	278	1 US-08-222-177A-52	Sequence 52, Appl
c 9	40.6	1.1	3757	2 US-09-016-366A-13	Sequence 13, Appl
c 10	40.6	1.1	3757	2 US-08-978-404B-19	Sequence 19, Appl
c 11	40.6	1.1	43795	3 US-08-742-185-101	Sequence 101, Appl
c 12	40.4	1.1	5134	2 US-08-310-912A-157	Sequence 157, Appl
c 13	40.4	1.1	5134	5 PCT-US95-04589-157	Sequence 157, Appl
c 14	40	1.1	22846	2 US-08-469-461-3	Sequence 3, Appl
c 15	40	1.1	22846	3 US-07-890-609-3	Sequence 3, Appl
c 16	38.8	1.1	177	2 US-08-829-961-3	Sequence 3, Appl
c 17	38.8	1.1	2618	4 US-09-488-671-17	Sequence 17, Appl
c 18	38.8	1.1	12141	4 US-09-488-671-10	Sequence 10, Appl
c 19	38	1.1	26700	1 US-08-472-217-1	Sequence 1, Appl
c 20	38	1.1	26700	2 US-08-488-199-5	Sequence 5, Appl
c 21	38	1.1	26700	3 US-08-760-534A-1	Sequence 1, Appl
c 22	37.6	1.0	2529	3 US-09-051-969A-5	Sequence 5, Appl
c 23	37.6	1.0	11236	1 US-07-853-913-1	Sequence 1, Appl
c 24	37.4	1.0	2775	1 US-08-149-096A-1	Sequence 1, Appl
c 25	37.2	1.0	223	1 US-08-222-177A-14	Sequence 14, Appl
c 26	37.2	1.0	3588	1 US-08-197-792-32	Sequence 32, Appl
c 27	37.2	1.0	3588	1 US-08-459-850-32	Sequence 32, Appl

c 28	37.2	1.0	3588	1 US-08-459-214-32	Sequence 32, Appl
c 29	37	1.0	265	1 US-08-222-177A-1	Sequence 1, Appl
c 30	37	1.0	1493	1 US-08-340-820-24	Sequence 24, Appl
c 31	37	1.0	1493	1 US-08-593-535-24	Sequence 24, Appl
c 32	36.6	1.0	6350	2 US-08-385-339A-9	Sequence 9, Appl
c 33	36.4	1.0	298	1 US-08-599-252-88	Sequence 88, Appl
c 34	36.4	1.0	298	5 PCT-US96-06352-88	Sequence 88, Appl
c 35	36.4	1.0	298	5 PCT-US96-06352-88	Sequence 88, Appl
c 36	36.4	1.0	600	1 US-08-599-252-104	Sequence 104, Appl
c 37	36.4	1.0	600	5 PCT-US96-06352-104	Sequence 104, Appl
c 38	36.4	1.0	600	5 PCT-US96-06352-104	Sequence 104, Appl
c 39	36.4	1.0	1700	5 PCT-US92-02091-1	Sequence 1, Appl
c 40	36.4	1.0	246240	2 US-08-724-394A-20	Sequence 20, Appl
c 41	36.4	1.0	246240	2 US-08-724-394A-21	Sequence 21, Appl
c 42	36.4	1.0	246240	2 US-08-724-394A-22	Sequence 22, Appl
c 43	36.2	1.0	360	4 US-08-118-200-12	Sequence 12, Appl
c 44	36.2	1.0	360	4 US-08-458-745-13	Sequence 13, Appl
c 45	36.2	1.0	2980	2 US-08-463-081B-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
; US-08-232-463-14

Qy 2086 tctatgatatatgtatgtatgaacataccaaagtatgcctctctctcttga 2145

APPLICANT: Stevens, Richard L.
TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
TITLE OF INVENTION: FIBRINOGEN
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
STREET: 600 Atlantic Avenue
CITY: Boston

CITY: BOSTON
STATE: MA
COUNTRY: U.S.A.
ZIP: 02210-2211
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/978,404B
FILING DATE: 25-NOV-97
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/032,354
FILING DATE: 04-DEC-1996

NAME: Plumer, Elizabeth R.
REGISTRATION NUMBER: 36, 637
REFERENCE/DOCKET NUMBER: B0801/7090
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
TELEX:

QY	2086	tctgtgctgtgctgtgctgtgtgtgagtgatacatgccaaagtatgcctctctctcttga	2145
Db	1321	TGCTGTGTGTGTGTGTGTGTGTGTGTGTATGTGTGTGTATGACTGGT	1380
QY	2146	cccatatttcadgaacttaa	2164
Db	1381	CCTCTCATTTGCACTCAA	1399

US-08-742-185-101/c
; Sequence 101, Application US/08742185
; Patent No. 6020476
; GENERAL INFORMATION:

APPLICANT: Page, David C.
APPLICANT: Reijo, Renee
APPLICANT: Saxena, Richa
APPLICANT: Hawkins, Trevor
APPLICANT: Reeve, Mary Pat
TITLE OF INVENTION: DAZ: A GENE FAMILY ASSOCIATED WITH AZOOSPERMIA
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:

APPLICANT: Resve, Mary Pat
TITLE OF INVENTION: DAY: A GENE FAMILY ASSOCIATED WITH
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: US
ZIP: 02173
COMPUTER READABLE FORM:

STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: US
ZIP: 02173
COMPUTER READABLE FORM:

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: July 31, 2001, 07:26:25 ; Search time 478.87 Seconds
(without alignments)
4696.773 Million cell updates/sec

Title: US-09-402-713A-6

Perfect score: 3582

Sequence: 1 acagaagaataagcaagtgc.....tgattctttgttacaacttt 3582

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq_0601.*
1: /SID88/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3580.4	100.0	3582	19 AAV62430	Prostate cancer an
2	2032	56.7	2037	19 AAV62427	Prostate cancer an
3	1772.4	49.5	1872	19 AAV62428	Prostate cancer an
4	1742.2	48.6	3112	21 AAA06687	Human immunogenic
C 5	1733.4	48.4	2426	21 AAA06689	Human immunogenic
C 6	1717.4	47.9	2229	21 AAA06688	Human immunogenic
7	725.4	20.3	820	19 AAV62429	Prostate cancer an
C 8	722.4	20.2	812	21 AAA06690	Human immunogenic
9	524	14.6	597	20 AAX37486	Human secreted pro
10	457.2	12.8	718	21 AAA06545	Human immunogenic
11	319.6	8.9	437	21 AAC06768	Human secreted pro

C	12	288.4	8.1	301	21	AAA06520	Human immunogenic
	13	247	6.9	283	21	AAA06468	Human immunogenic
	14	235.4	6.6	359	20	AAZ33445	Human prostate can
C	15	133	3.7	936	22	AAF58252	Oligonucleotide D1
C	16	133	3.7	936	22	AAF58254	Oligonucleotide D1
C	17	133	3.7	936	22	AAF58257	Oligonucleotide D1
C	18	133	3.7	936	22	AAF58259	Oligonucleotide D2
C	19	133	3.7	936	22	AAF58262	Oligonucleotide D2
C	20	133	3.7	936	22	AAF58255	Oligonucleotide D1
	21	132.6	3.7	936	22	AAF58255	Oligonucleotide D1
	22	132.6	3.7	936	22	AAF58254	Oligonucleotide D1
	23	132.6	3.7	936	22	AAF58257	Oligonucleotide D1
	24	132.6	3.7	936	22	AAF58259	Oligonucleotide D2
	25	132.6	3.7	936	22	AAF58262	Oligonucleotide D2
	26	132.6	3.7	936	22	AAF58255	Oligonucleotide D1
	27	121	3.4	123	19	AAV33791	Prostate cancer an
	28	52.2	1.5	244	22	AAF58238	Oligonucleotide D1
C	29	51.4	1.4	244	22	AAF58238	Oligonucleotide D1
	30	44.2	1.2	3425	15	AAV57072	AGE-modified DNA I
C	31	43	1.2	10968	19	AAV17789	Tomato Prf genomic
C	32	42.4	1.2	3282	21	AAAS3887	Murine licby recep
C	33	41.6	1.2	1287	21	AAC59013	Human secreted pro
C	34	41.6	1.2	10409	19	AAV42558	Mouse dectin-2 gen
C	35	41.4	1.2	5475	19	AAV17777	Tomato Prf cDNA.
	36	40.8	1.1	319	20	AAV90110	EST clone DG329.
	37	40.6	1.1	278	18	AAV65703	Polymorphic repeat
	38	40.6	1.1	3757	19	AAV44332	Murine mMCP-6 zymo
	39	40.6	1.1	3757	19	AAV42708	Mouse mast cell pr
C	40	40.6	1.1	43795	21	AAZ92583	Human DAZ genomic
	41	40.2	1.1	1729	20	AAX86275	DNA encoding a hum
C	42	40	1.1	796	20	AAX37484	Human secreted pro
	43	40	1.1	3710	18	AAV74716	Staphylococcus aur
C	44	39.4	1.1	1660	22	AAF81244	Anthranilate synth
C	45	39.2	1.1	541	21	AAA74269	Loiblolly pine SSR

ALIGNMENTS

RESULT 1
AAV62430
ID AAV62430 standard; cDNA; 3582 BP.
XX
AC AAV62430;
XX
DT 30-DEC-1998 (first entry)
XX
DE Prostate cancer antigen (PCA3) wild-type cDNA.
XX
KW Prostate cancer antigen cDNA; PCA3; prostatic cancer;
KW PC; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 401..556
FT /*tag= a
FT /product= "PCA3 protein"
FT polyA_signal 983..987
FT /*tag= b
FT polyA_signal 2041..2046
FT /*tag= c
FT polyA_signal 2597..2602
FT polyA_signal 3494..3496
FT /*tag= d
FT /*tag= e
PN WO9845420-A1.
XX
PD 15-OCT-1998.
XX
PF 09-APR-1998; 98WO-CA00346.
XX


```
Qy 1621 tcttacttcatgcaagaagggacacatatagagattcatcatcaccatgagacagcaata 1680
Db 1621 tcttacttcatgcaagaagggacacatatagagattcatcatcaccatgagacagcaata 1680
Qy 1681 ctaaaagtgttaattgtattataaagatttagataaataataagaatgcaagagccacaga 1740
Db 1681 ctaaaagtgttaattgtattataaagatttagataaataataagaatgcaagagccacaga 1740
Qy 1741 gggaaatgtttatggggcagctttgtaagcctgggatgtgaagcaaaagcagggaacctca 1800
Db 1741 gggaaatgtttatggggcagctttgtaagcctgggatgtgaagcaaaagcagggaacctca 1800
Qy 1801 tagtatcttatataataacttacttctctctctctctctctctctctctctctctctctct 1860
Db 1801 tagtatcttatataataacttacttctctctctctctctctctctctctctctctctctct 1860
Qy 1861 cacagaattcatgcaagtgaatacccccaaggtaacctttatccatttcaatggtgagtc 1920
Db 1861 cacagaattcatgcaagtgaatacccccaaggtaacctttatccatttcaatggtgagtc 1920
Qy 1921 gctttagaattttggcaaatcactgctcacttacttacttacttacttacttacttactt 1980
Db 1921 gctttagaattttggcaaatcactgctcacttacttacttacttacttacttacttactt 1980
Qy 1981 ttgtagttaattgaaagaaatagggcactcttgtgagccacttttagggttcactcctggc 2040
Db 1981 ttgtagttaattgaaagaaatagggcactcttgtgagccacttttagggttcactcctggc 2040
Qy 2041 aataaagaatttacaaagagctactcaggaccagttgttaagagctctgtgtgtgtgt 2100
Db 2041 aataaagaatttacaaagagctactcaggaccagttgttaagagctctgtgtgtgtgt 2100
Qy 2101 gtgtgtgtgtgagtgatcagccaaagtgtccctctctctctctctctctctctctctct 2160
Db 2101 gtgtgtgtgtgagtgatcagccaaagtgtccctctctctctctctctctctctctctct 2160
Qy 2161 ttaaaacaagcatgttttcaaatggcactatgagctgccaatgtatgtatcaccaccatat 2220
Db 2161 ttaaaacaagcatgttttcaaatggcactatgagctgccaatgtatgtatcaccaccatat 2220
Qy 2221 ctcatattctccagtaaatgtataataatgtatcctctgtttaaataaaaaaagtgtgac 2280
Db 2221 ctcatattctccagtaaatgtataataatgtatcctctgtttaaataaaaaaagtgtgac 2280
Qy 2281 ttcacaaagcagctggaaatggacacacacacacacacacacacacacacacacacacac 2340
Db 2281 ttcacaaagcagctggaaatggacacacacacacacacacacacacacacacacacacac 2340
Qy 2341 gctcacacactgctgacatatattgttagaagcactcgcattgtgggttctctcttaagc 2400
Db 2341 gctcacacactgctgacatatattgttagaagcactcgcattgtgggttctctcttaagc 2400
Qy 2401 aaaaactctgattaggtctcagctgggctgtgcatcaggcggtttgagaaaatttcaa 2460
Db 2401 aaaaactctgattaggtctcagctgggctgtgcatcaggcggtttgagaaaatttcaa 2460
Qy 2461 ttctcagcagaagcagaatttgaattccctcactctcttttagaatacatttaccagggttg 2520
Db 2461 ttctcagcagaagcagaatttgaattccctcactctcttttagaatacatttaccagggttg 2520
Qy 2521 gagaggattcagacagctcagggtgtttcactaatgtctgaaactctgtccctctttg 2580
Db 2521 gagaggattcagacagctcagggtgtttcactaatgtctgaaactctgtccctctttg 2580
Qy 2581 tgttcattgtagtcccaataataatgttatcttttgaactgatgctcatagaggagagata 2640
Db 2581 tgttcattgtagtcccaataataatgttatcttttgaactgatgctcatagaggagagata 2640
Qy 2641 taagaacctctgagtgatatacaacattagggattcaaaagaaatatttagatttaagctcaca 2700
Db 2641 taagaacctctgagtgatatacaacattagggattcaaaagaaatatttagatttaagctcaca 2700
```

```
Qy 2701 ctgggtcaaaaagaaaccaagatacacaagaactctgagctgtcatctccccatctctgtga 2760
Db 2701 ctgggtcaaaaagaaaccaagatacacaagaactctgagctgtcatctccccatctctgtga 2760
Qy 2761 gccacaaaccaacagcaggacccaacgcgcatgtctgagatactctaaatacaaggaaacagtg 2820
Db 2761 gccacaaaccaacagcaggacccaacgcgcatgtctgagatactctaaatacaaggaaacagtg 2820
Qy 2821 tcatgagtgtaattctctctctctctctctctctctctctctctctctctctctctctct 2880
Db 2821 tcatgagtgtaattctctctctctctctctctctctctctctctctctctctctctctct 2880
Qy 2881 gacacatatagctctcagctctgtcttctccacgacttttattcttctccacacatcgc 2940
Db 2881 gacacatatagctctcagctctgtcttctccacgacttttattcttctccacacatcgc 2940
Qy 2941 ttaccaaactctctctctctctctctctgtgtgttggaacttccccacaagaatttcaacgact 3000
Db 2941 ttaccaaactctctctctctctctctgtgtgttggaacttccccacaagaatttcaacgact 3000
Qy 3001 caaagtctttctctccatccccacacactaacctggaattgcttagacccctattcttataa 3060
Db 3001 caaagtctttctctccatccccacacactaacctggaattgcttagacccctattcttataa 3060
Qy 3061 ttccaatagatgctgctctatgggctaattattgcttttagatgaacatttagatttaaag 3120
Db 3061 ttccaatagatgctgctctatgggctaattattgcttttagatgaacatttagatttaaag 3120
Qy 3121 tetaagaggttcaaaaataccaaactattatctctctcttcttcttctccactccctgctct 3180
Db 3121 tetaagaggttcaaaaataccaaactattatctctctcttcttcttctccactccctgctct 3180
Qy 3181 cccatatattactgattgactgaacaggatgtccccagatgcccagtcacaaatgagaacc 3240
Db 3181 cccatatattactgattgactgaacaggatgtccccagatgcccagtcacaaatgagaacc 3240
Qy 3241 cagtggtcctctgtggtcatgcatgcaagactgctgaagccagaggatgactgattacg 3300
Db 3241 cagtggtcctctgtggtcatgcatgcaagactgctgaagccagaggatgactgattacg 3300
Qy 3301 cctcatgtggtggaggggacccactcctggcctctgtgtgtgtcagggacgaacacctgaga 3360
Db 3301 cctcatgtggtggaggggacccactcctggcctctgtgtgtgtcagggacgaacacctgaga 3360
Qy 3361 tgcctcctgctctcagtgctctctcactcctcccttcttaataagatccatagaaatttg 3420
Db 3361 tgcctcctgctctcagtgctctctcactcctcccttcttaataagatccatagaaatttg 3420
Qy 3421 ctacatttgagaattcccaattaggaaactcacaatgttttatctgcccctatcaatttttaa 3480
Db 3421 ctacatttgagaattcccaattaggaaactcacaatgttttatctgcccctatcaatttttaa 3480
Qy 3481 acttctgaaaaataaagtttttccaaaatctgctctgttaaattactttttcttaccagtg 3540
Db 3481 acttctgaaaaataaagtttttccaaaatctgctctgttaaattactttttcttaccagtg 3540
Qy 3541 tcttgagcactatatacaactttgattcttctgttacaacttt 3582
Db 3541 tcttgagcactatatacaactttgattcttctgttacaacttt 3582
RESULT 2
AAV62427
ID AAV62427 standard; cDNA; 2037 BP.
XX
AC AAV62427;
XX
DT 30-DEC-1998 (first entry)
XX
DE Prostate cancer antigen (PCA3) cDNA splice variant 1.
XX
KW Prostate cancer antigen cDNA splice variant 1; PCA3; prostatic cancer;
PC; ds.
```

XX Homo sapiens.
OS
XX Key Location/Qualifiers
FH CDS 379..534
FT /*tag= a
FT /product= "PCA3 protein variant 1"
FT polyA_signal 2019..2024
FT /*tag= b
XX
XX WO9845420-A1.
XX
XX PN
XX
XX PD 15-OCT-1998.
XX
XX PF 09-APR-1998; 98WO-CA00346.
XX PR 10-APR-1997; 97US-0041836.
XX
XX PA (DIAG-) DIAGNOCURE INC.
XX
XX PI Bussemakers MJG;
XX
XX WPI: 1998-568347/48.
XX P-PSDB; AAW79736.
XX
XX PT New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
XX prevention and treatment of prostatic cancer
XX
XX PS Claim 3; Fig 2B-2J; 111pp; English.
XX
XX The present sequence represents the prostate cancer antigen (PCA3)
CC cDNA splice variant 1 sequence comprising of exons 1, 2, 3, 4a and
CC 4b of the PCA3 gene. The PCA3 cDNA splice variant 1 sequence,
CC isolated from a human primary prostatic tumour tissue cDNA library,
CC was found in approximately 5% of the cDNA clones isolated. The
CC invention claims for PCA3 cDNA variants and the proteins they encode.
CC The invention also claims for antibodies against PCA3 protein. The
CC antibodies are claimed to be useful for detecting PCA3 protein in
CC immunoassay tests, for diagnosing, assessing and prognosing of
CC prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin
CC or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed
CC to be useful for treating PC, while determining elevated levels of
CC PCA3 (as RNA or protein) is useful for detecting a predisposition
CC to development of PC, e.g. in prenatal tests. Detecting PCA3 protein
CC allows differentiation between malignant and benign prostatic disease,
CC and the level of PCA3 expression allows correlation with the grade of
CC tumour. PCA3 protein and its fragments are also claimed to be useful
CC in vaccines for preventing PC; in drug screens for identifying
CC specific (antagonists (potentially useful therapeutically) and for
CC studying protein-DNA interactions.
XX
XX Sequence 2037 BP; 622 A; 426 C; 406 G; 575 T; 8 other;

Query Match 56.7%; Score 2032; DB 19; Length 2037;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 2029; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 23 agaagctgcgcacagaaacagagggagattgtgtgctgcagccgagggagaccag 82
DB 1 agaagctgcgcacagaaacagagggagattgtgtgctgcagccgagggagaccag 60
QY 83 gaagactctgctggtgggaagacctgatgatcacagaggaataacacacatacttag 142
DB 61 gaagactctgctggtgggaagacctgatgatcacagaggaataacacacatacttag 120
QY 143 tgtttcaatgaacacacagataaataagtgagagctagctcgtgtgagctcctcagt 202
DB 121 tgtttcaatgaacacacagataaataagtgagagctagctcgtgtgagctcctcagt 180
QY 203 gacacagggctggatcaccatcgagggcacttctgagctactcagtgacgacaaagaaga 262
DB 181 gacacagggctggatcaccatcgagggcacttctgagctactcagtgacgacaaagaaga 240

QY 263 ctacagacatctcaatggcaggggtgagaaataagaaaggctgctgactttaccatctga 322
DB 241 ctacagacatctcaatggcaggggtgagaaataagaaaggctgctgactttaccatctga 300
QY 323 ggcacacatctgctgaaatggagataataacatcactagaaacagcagagatgacaata 382
DB 301 ggcacacatctgctgaaatggagataataacatcactagaaacagcagagatgacaata 360
QY 383 taatgtctaagttagtgacatgtttttgacatttcagcccttttaataatccacacaca 442
DB 361 taatgtctaagttagtgacatgtttttgacatttcagcccttttaataatccacacaca 420
QY 443 caggaagcacaagaagacacagagatccctgggagaaatgcccgccgcgcacatcttggg 502
DB 421 caggaagcacaagaagacacagagatccctgggagaaatgcccgccgcgcacatcttggg 480
QY 503 tcacgatgagcctgcctctgtgctgtgtcccgcttgtgagggaaaggacattagaaaaatg 562
DB 481 tcacgatgagcctgcctctgtgctgtgtcccgcttgtgagggaaaggacattagaaaaatg 540
QY 563 aattgatgtgttccttaaaaggatgggcagagaaacagatccctgttggatattattttg 622
DB 541 aattgatgtgttccttaaaaggatgggcagagaaacagatccctgttggatattattttg 600
QY 623 aacgggattacagatttgaatgaagtacaaagtgcattaccacatgagaggaataaca 682
DB 601 aacgggattacagatttgaatgaagtacaaagtgcattaccacatgagaggaataaca 660
QY 683 gacgagaaatcttgatggcttcacaaagacatgcacaaacaaatggaaatctgtatg 742
DB 661 gacgagaaatcttgatggcttcacaaagacatgcacaaacaaatggaaatctgtatg 720
QY 743 acatgaggcagcaagctggggagagataaacacaggggcagagggctcaggattctggcc 802
DB 721 acatgaggcagcaagctggggagagataaacacaggggcagagggctcaggattctggcc 780
QY 803 ctgctgcctaaactgtcgcttataaccacaaatcatttcatttcttaacccctcaaaaaca 862
DB 781 ctgctgcctaaactgtcgcttataaccacaaatcatttcatttcttaacccctcaaaaaca 840
QY 863 agctgtgtaatatctgatctctacaggttcctctctggtggcccaacattctccatatcca 922
DB 841 agctgtgtaatatctgatctctacaggttcctctctggtggcccaacattctccatatcca 900
QY 923 gccacactatttttaataatttagttccagatctgtactgtgacctttctacactgtag 982
DB 901 gccacactatttttaataatttagttccagatctgtactgtgacctttctacactgtag 960
QY 983 aataacattactcattttgttcaaaagacccttcgtgtgtgctgcctaaatagttagctgact 1042
DB 961 aataacattactcattttgttcaaaagacccttcgtgtgtgctgcctaaatagttagctgact 1020
QY 1043 gtttttctaaaggagtgttctggccagggggtatctgtgaacagggtgggaagcatctcaaa 1102
DB 1021 gtttttctaaaggagtgttctggccagggggtatctgtgaacagggtgggaagcatctcaaa 1080
QY 1103 gatcttcaggggttatacttactagcacacagcatgatcattacaggagtgaattatcta 1162
DB 1081 gatcttcaggggttatacttactagcacacagcatgatcattacaggagtgaattatcta 1140
QY 1163 atcaacatcactcagtgctttgtcccatagctgaaatcatttcccacttttggccca 1222
DB 1141 atcaacatcactcagtgctttgtcccatagctgaaatcatttcccacttttggccca 1200
QY 1223 ttctcaagacctcaaaaatgtcattcatttaataatcacaggatttaactttttttaacc 1282
DB 1201 ttctcaagacctcaaaaatgtcattcatttaataatcacaggatttaactttttttaacc 1260
QY 1283 tggagaattcaatgttacatgcagctatgggaatttaattacatatttctgttttccagt 1342
DB 1261 tggagaattcaatgttacatgcagctatgggaatttaattacatatttctgttttccagt 1320

QY 821 gttcataaccaaatcatcttcatatatttctaacctcctcaaaacaaagctgttgttaatatctga 880
Db 634 gttcataaccaaatcatcttcatatatttctaacctcctcaaaacaaagctgttgttaatatctga 693
QY 881 tctctacggttctcttctgggcccacaaatctctcaatatataccagccacacactcattttta 940
Db 694 tctctacggttctcttctgggcccacaaatctctcaatatataccagccacacactcattttta 753
QY 941 atttagttccagatctgactgtgacctttctacactgtagaataaacattactcatttt 1000
Db 754 atttagttccagatctgactgtgacctttctacactgtagaataaacattactcatttt 813
QY 1001 gttcaaaagacctctgttctgtcgttaatatgtactgttcttcttaaggagtgt 1060
Db 814 gttcaaaagacctctgttctgtcgttaatatgtactgttcttcttaaggagtgt 873
QY 1061 tctggcccaggggatctgtgaacagctgggaagcattctcaagatcttccagggttata 1120
Db 874 tctggcccaggggatctgtgaacagctgggaagcattctcaagatcttccagggttata 933
QY 1121 ctctactgacacagcatgatcattacgagtgaaattatctcaatcaacatcatcctcagt 1180
Db 934 ctctactgacacagcatgatcattacgagtgaaattatctcaatcaacatcatcctcagt 993
QY 1181 gtcttggccataactgaaattcattcccaacttttggcccattctcaagacctcaaaa 1240
Db 994 gtcttggccataactgaaattcattcccaacttttggcccattctcaagacctcaaaa 1053
QY 1241 gtcattccatttaatacacaggattaaacttttttttaacctgggaagattcaatgtta 1300
Db 1054 gtcattccatttaatacacaggattaaacttttttttaacctgggaagattcaatgtta 1113
QY 1301 catcgagctatgggaatttaattacataattttgtttccagtgcaagatgactaagtc 1360
Db 1114 catcgagctatgggaatttaattacataattttgtttccagtgcaagatgactaagtc 1173
QY 1361 ttatccctccctctgttggatttttttccagtgataaagttaaaacttagccttgt 1420
Db 1174 ttatccctccctctgttggatttttttccagtgataaagttaaaacttagccttgt 1233
QY 1421 actgagctgtatcacagcagcctccctccctccctccctccctccctccctccctcc 1480
Db 1234 actgagctgtatcacagcagcctccctccctccctccctccctccctccctccctcc 1293
QY 1481 caacctccctccctccctccctccctccctccctccctccctccctccctccctccctcc 1540
Db 1294 caacctccctccctccctccctccctccctccctccctccctccctccctccctccctcc 1353
QY 1541 atacattatctctgctggaagctcttctctgtctctttaaactagaatgatgtaa 1600
Db 1354 atacatttctctgctggaagctcttctctgtctctttaaactagaatgatgtaa 1413
QY 1601 agtttgaataagttgactattctactctcatgcaagaaggacacatatgagattcacc 1660
Db 1414 agtttgaataagttgactattctactctcatgcaagaaggacacatatgagattcacc 1473
QY 1661 atcacatgagacgcaaaactactaaaagtgtatttgaattagaagtttagataaata 1720
Db 1474 atcacatgagacgcaaaactactaaaagtgtatttgaattagaagtttagataaata 1533
QY 1721 tgaattgcaagagccacagaggaaatgtttatggggcacgttttgaagcctggagtga 1780
Db 1534 tgaattgcaagagccacagaggaaatgtttatggggcacgttttgaagcctggagtga 1593
QY 1781 agcaagagcaggaaacctcatagattcttatataataacttacttctctatctctacc 1840
Db 1594 agmaaggcaggaaacctcatagattcttatataataacttacttctctatctctacc 1653
QY 1841 acaatatcaaacagcttttccagaaattctagcagtgcaaatcccccaaggtaaccttt 1900
Db 1654 acaatatcaaacagcttttccagaaattctagcagtgcaaatcccccaaggtaaccttt 1713

QY 1901 atocatttcagtgagtgagtcgcttttagaatttttggcaaatcactatggtcacttatctca 1960
Db 1714 atocatttcagtgagtgagtcgcttttagaatttttggcaaatcactatggtcacttatctca 1773
QY 1961 actttgagatggtttgtctctgttagttaattgaaagaataaggcactcttctgagacca 2020
Db 1774 actttgagatggtttgtctctgttagttaattgaaagaataaggcactcttctgagacca 1833
QY 2021 ctttaggggttactcctcctggaataaaagaatttacaaga 2059
Db 1834 ctttaggggttactcctcctggaataaaagaatttacaaga 1872

RESULT 4

AAA06687
ID AAA06687 standard; cDNA; 3112 BP.

XX AC AAA06687;
XX DT 13-JUN-2000 (first entry)
XX DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:468.
XX DE Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW immunogenic; cytostatic; vaccine; ss.
XX OS Homo sapiens.
XX PN WO200004149-A2.
XX PD 27-JAN-2000.
XX PF 14-JUL-1999; 99WO-US15838.
XX PR 14-JUL-1998; 98US-0115453.
XX PR 14-JUL-1998; 98US-0115134.
XX PR 23-SEP-1998; 98US-0159812.
XX PR 23-SEP-1998; 98US-0159822.
XX PR 15-JAN-1999; 99US-0232149.
XX PR 15-JAN-1999; 99US-0232880.
XX PR 09-APR-1999; 99US-0288946.
XX PA (CORI-) CORIXA CORP.

PI Dillon DC, Harlocker SL, Yugu J, Xu J, Mitcham JL;
XX WPI; 2000-171268/15.
XX PT New polypeptide useful for treating and diagnosing prostate cancer
XX comprises an immunogenic portion of prostate tumor protein .
XX PS Claim 1; Page 259-260; 263pp; English.
XX CC The present invention describes isolated polypeptides, comprising an
XX immunogenic portion of a prostate tumour protein (PTP). The polypeptides
XX and polynucleotides encoding them have cytostatic activity and can be
XX used in vaccines and in gene therapy. The polypeptides and
XX polynucleotides encoding them, antigen presenting cells which express
XX the polypeptides, antibodies against the polypeptides and vaccines
XX comprising them can be used for inhibiting the development of prostate
XX cancer in a patient. The polypeptides can be used to generate antibodies
XX or anti-idiotypic antibodies for passive immuno therapy. A portion of
XX the polynucleotides encoding the polypeptides can be used as a probe or
XX to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
XX CC AAY82000 to AAY82020 represent sequences used in the exemplification of
XX the present invention.

XX SQ Sequence 3112 BP; 975 A; 587 C; 624 G; 926 T; 0 other;
Query Match 48.6%; Score 1742.2; DB 21; Length 3112;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1780; Conservative 0; Mismatches 8; Indels 3; Gaps 3;

PR 23-SEP-1998; 98US-0159822.
 PR 15-JAN-1999; 99US-0232149.
 PR 15-JAN-1999; 99US-0232880.
 PR 09-APR-1999; 99US-0288946.
 XX
 XX (CORI-) CORIXA CORP.
 PA
 XX
 PI Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;
 DR
 XX
 DR WPI; 2000-171268/15.
 XX
 XX
 PT New polypeptide useful for treating and diagnosing prostate cancer
 PT comprises an immunogenic portion of prostate tumor protein -
 XX
 XX
 PS Claim 1; Page 261-262; 263pp; English.
 XX
 CC The present invention describes isolated polypeptides, comprising an
 CC immunogenic portion of a prostate tumour protein (PrP). The polypeptides
 CC and polynucleotides encoding them have cytostatic activity and can be
 CC used in vaccines and in gene therapy. The polypeptides and
 CC polynucleotides encoding them, antigen presenting cells which express
 CC the polypeptides, antibodies against the polypeptides and vaccines
 CC comprising them can be used for inhibiting the development of prostate
 CC cancer in a patient. The polypeptides can be used to generate antibodies
 CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
 CC the polynucleotides encoding the polypeptides can be used as a probe or
 CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
 CC AAY82000 to AAY82020 represent sequences used in the exemplification of
 CC the present invention.
 XX
 XX Sequence 2426 BP; 717 A; 476 C; 548 G; 685 T; 0 other;

Query Match 48.4%; Score 1733.4; DB 21; Length 2426;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1767; Conservative 0; Mismatches 1; Indels 3; Gaps 3;

QY 285 ggtgaaataagaaggctgactgtaccatctgagccacacatctgctgaaatgg 344
 DB 1769 GGTGAAATAAGAAAGCTGCTGACTTTACCATCTGAGGCCACACATCTGCTGAATGG 1710
 QY 345 agataataacatcactagaacagcaagatgacaataataatgtcttaagttagcatgt 404
 DB 1709 AGATAATTAAACATCAGTAGAAACAGCAAGATGACATAATGTCTTAAAGTAGTACATGT 1650
 QY 405 tttgcacatttcacgccccttaataatccacacacaggaagcacaaaaggagcac 464
 DB 1649 TTTTGCACATTTCCAGCCCTTTAAATATCCACACACAGGAGCACAAGGAAGCAC 1590
 QY 465 agagatccctgggagaaatgcccgccgcacatcttgggtcatcgatgagcctcgccctgt 524
 DB 1589 AGAGATCCCTGGGAGAAATGCCGGCCGCATCTTGGGTATCGATGAGCCTCGCCCTGT 1530
 QY 525 gctgtcccgctgtgaggaaggacattagaaatgaattgatgttcccttaagga 584
 DB 1529 GCTGTGCTCCGCTGTGAGGAAGGACATTAGAAATGAATGATGTGTCTTAAAGGA 1470
 QY 585 tgggcaggaaacacagatcctgttggatatttatttgaacgggattacagatttgaat 644
 DB 1469 TGGCAGGAAACACAGATCTGTGTGGATATTATTTGAACGGATTACAGATTTCGAAT 1410
 QY 645 gaagtcaaaagttagcatataccaatgagagaaacacagacagaaaaatcttgatggctt 704
 DB 1409 GAAGTCACAAAGTGACATTTACCAATGAGAGGAAACACAGACGAGAAATCTTGTATGGCTT 1350
 QY 705 cacaagacatgcaacaaataatgaatgactgtgatgacatgagcagcaagctgggg 764
 DB 1349 CACAAGCATGTCAACAAACAAATGAAATGACTGTGTATGACATGAGCGACGCCAAGCTGGGG 1290
 QY 765 agagataaccacggggcagaggtcagatcttctgcccctgtgcctaaactgtgcgttc 824
 DB 1289 AGGAGATACCACGGGCGAGGGTCAGATCTTGGCCCTGTGCTTAACTTAACTTGGCTTC 1230

QY 825 ataacaataatcttcaatatttcaaacctcaaaacaaagctgtgtgataatatctatc 884
 DB 1229 ATAACCAATCATTTTCAATTTCTAACCCCTCAAAACAAAGCTGTGTAAATATCTGATCTC 1170
 QY 885 tacggttcccttctggggcccaacattctccatatatccagccacactcatttttaatttt 944
 DB 1169 TAGGGTTCTTCTGGGCCCAACATTTCTCCATATATCAGCCACACATTTTAAATATTT 1110
 QY 945 agtccacagatctgtactgtgacctttcacactgtagaataacattcattttgttc 1004
 DB 1109 AGTCCACAGATCTGTACTGTGACCTTTCACACTGTAGAATAACATTAATCTCAATTTTGTTC 1050
 QY 1005 aaagacccctcgtgtgtgctgcctaaatagtactgactgttttccctaaaggagtgcttg 1064
 DB 1049 AAAGACCCCTTCTGTGTGCTGCCTAATATGCTAGCTGACTGTTTTTCTTAAGGAGTGTCTG 990
 QY 1065 gccaggggatctgtgaaacagcgtgggaagcatctcaagatctttccaggggttatactta 1124
 DB 989 GCCCAGGGGATCTGTGAACAGGCTGGGAAGCATCTCAAGATCTTTCCAGGGTTATATCTTA 930
 QY 1125 ctagcacacagcatgatcattacgagtgaaattatctaatcaacatcctcctcagtgct 1184
 DB 929 CTAGCACACAGCATGATCAATACGGAGTGAATTAATCAACATCACTCCCTCAGTGTCT 870
 QY 1185 ttgcccactactgaaattcatttcccaacttttgcaccttttgcaccttcccaagacacacaaatgtca 1244
 DB 869 TTGCCCATACTGAAATTCATTTCCCACTTTTGTGCCCATTTCTCAAGACCTCAAAATGTCA 810
 QY 1245 ttccattaatatacacaggaattaaacttttttttaacctgggaagaattcaattgttacatg 1304
 DB 809 TTCCATTAATATCACAGGATTAACTTTTTTTTTAACCTGGGAAGAAATCAATGTTTACATG 750
 QY 1305 cagctatgggaatttaattacatatatttttccagtgcaaaagatgactaaagctctta 1364
 DB 749 CAGCTATGGGAATTAATTAATTAATTAATTTTCCAGTCAAAAGATGACTAAGTCCCTTTA 690
 QY 1365 tccctccccttgttggatttttttccagtgataaaagttaaaatgcttagcctgtactg 1424
 DB 689 TCCCTCCCTTTGTTGATTTTTTTTCCAGTATAAAAGTTAAATGCTTAGCCTTGTACTG 630
 QY 1425 aggctgtatacag-cacagcctctcccacatccctccagccttatctgtcacaacataa 1483
 DB 629 AGGCTGTATACAGCCACAGCCCTTCCCCATCCCTCCAGCTTATCTGTCTATCACCACAA 570
 QY 1484 cccctcccacacccaaacaaatcctaactgttaacttcttgaacatgtcaggaacata 1543
 DB 569 CCCCTCCCAT-CCACCTAAACAAAATCTAACTGTGAATTTCTTGAACATGTGAGG-CATA 512
 QY 1544 cattatcccttgcctgagaagctcttccctgtctcttaaatctagaatgatgtaagt 1603
 DB 511 CATTAATCTCTGCTGAGAAAGCTCTTCTTGTCTCTTTAAATCTAGAATGATGTAAGT 452
 QY 1604 ttgtaataagtgactatcttactcaagaagggacacatagattcattcactc 1663
 DB 451 TTGTAATAAGTTGACTATCTTCTCATGAAGAAGGACACATATGAGATTCATCATC 392
 QY 1664 acatgagacagcaaatcactaaaaagtgaattattgattataagagtttagataaataatga 1723
 DB 391 ACATGAGACAGCAAAATACTAAAAGTGAATTTGATTATAAGAGTTTAGATAAATATATGA 332
 QY 1724 aatgaagagccacagaggaatgtttatggggcagcttgttaagcctgggagtggaagc 1783
 DB 331 AATGCAAGAGCCACAGAGGGAATGTTTATGGGGCAGCTTTGTAAGCCTGGGATGTGAAGC 272
 QY 1784 aaagcaggaagacctcatgatcttatataataataacttcttctcctcctcactcaca 1843
 DB 271 AAAGGCGAGGAACCTCATAGTATCTTATATAATATTAATTAATCTTCTATCTATCTATCA 212
 QY 1844 atatccacaagaacttttccacagaattcatcgagtgcaaatcccccaaggttaacacttate 1903
 DB 211 ATATCCACAAGCTTTTCCACAGAATTCATGAGTGCAATCCCAAGGTAACCTTTATC 152
 QY 1904 cattcatggtgagtcgctttagaattttggcaaatcatctggtcacttatctcaact 1963

|||||
Db 151 CATTTTCATGGTGGTGGCGCTTTAGAAATTTGGCAATCATACTGGTCACTTATCTCAACT 92
Qy 1964 ttgaagatgttttgcctttagttaaattgaagaataggcactctttagcactt 2023
Db 91 TTGAGATGTTGTTGCTTGTAGTTAAATTGAAGAAATAGGGCACTCTTGTGAGCCACTT 32
Qy 2024 tagggctcactctggcaataagaatttac 2054
Db 31 TAGGGTTCACTCTCTGGCAATAAAGAAATTAC 1
RESULT 6
ID AAA06688/c
XX AAA06688 standard; cDNA; 2229 BP.
AC AAA06688;
DT 13-JUN-2000 (first entry)
XX Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:469.
DE Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW immunogenic; cytostatic; vaccine; ss.
XX Homo sapiens.
XX WO200004149-A2.
PN 27-JAN-2000.
PD 14-JUL-1999; 99WO-US15838.
PF 14-JUL-1998; 98US-0115453.
PR 14-JUL-1998; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX (CORI-) CORIXA CORP.
XX Dillon DC, Harlocker SL, Yugu J, Xu J, Mitcham JL;
PI WPI; 2000-171268/15.
XX New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumor protein -
PS Claim 1; Page 260-261; 263pp; English.
XX The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate antibodies
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the polynucleotides encoding the polypeptides can be used as a probe or
CC to modulate the expression of the polypeptides; AAA06241 to AAA06691 and
CC AAA82000 to AAA82020 represent sequences used in the exemplification of
XX the present invention.
XX Sequence 2229 BP; 654 A; 447 C; 481 G; 647 T; 0 other;

Query Match 47.9%; Score 1717.4; DB 21; Length 2229;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1773; Conservative 0; Mismatches 1; Indels 5; Gaps 5;

Qy 285 ggtgagaaataagaagctgctgactttaccatctgagggccacacatctgctgaaatgg 344
Db 1775 GGTGAGAAATAAGAAAGGCTGCTGACTTTACCATCTGAGGCCACACATCTGCTGAATGG 1716
Qy 345 agataataacatcactcagaacacagagatgacaataataatgcttaagtagtagcatgt 404
Db 1715 AGATAATTAACATCACTAGAAAACAGCAAGATGACAATAATTAATGTCTTAAGTAGTGCATGT 1656
Qy 405 ttttgacatttccagcccttttaaatccacacacacagagaagcaaaaggaagcac 464
Db 1655 TTTTGCACATTTCCAGCCCTTTAAATATCCACACACAGGAGACACAAAAGGAGACAC 1596
Qy 465 agagatccctgggagaaatgccggcccatcttgggtcatcgatgagcctcgccctgt 524
Db 1595 AGAGATCCCTGGGAGAAATGCCGCCGCATCTTTGGGTATCGATGAGCCCTGCGCCTGT 1536
Qy 525 gctgtgctccgcttgtgaggggaagacattagaaaaatgaattgattgatttcccttaaaaga 584
Db 1535 GCCTGTGCTCCGCTTGTGAGGGAAGGACATTTAGAAAATGAATTTGATGTCTTCCCTTAAGGA 1476
Qy 585 tggcagagaaaaacagatcctgttggatatttttaaacgggattacagatttgaat 644
Db 1475 TGGGAGGAAAACAGATCCTGTGTGGATATTTATTGAACCGGATACAGATTTGAAT 1416
Qy 645 gaagtcaaaaagtgcagcattaccatgcagagaaacagacagagaaaaacttgcaggtt 704
Db 1415 GAAGTCACAAAGTGAGCATTTACCAATGAGGGAACACACGAGAAAATCTTTGATGGCTT 1356
Qy 705 cacaagacatgcacaaacaaaatggaatactgtgatgcacatgagggcagcagcagctgggg 764
Db 1355 CACAAGACATGCACAAACAAAATGGAATCTGTGATGACATGAGCGACGCCAAGCTGGGG 1296
Qy 765 agagataaacaacggggcagaggtcaggattctggccctgctgctaaactgtgcgttc 824
Db 1295 AGGAGATAACCCACGGGGCAGAGGGTCAGGATTTCTGGCCCTGCTGCTAACTTAACTGCTTC 1236
Qy 825 ataaccaaatcatttcataatttctaacctcacaagctgtgtgtaatactcgtatctc 884
Db 1235 ATAACCAAAATCATTTTCATATTTCTAACCTCAAAAACAAAGCTGTGTGTAATATCTGATCTC 1176
Qy 885 tacggtcctctctggggcccaacatctccatatataccagccacacactcttttaatat 944
Db 1175 TAGGGTTCTTCTGGGCCCAACATTTCTCCATATATCCAGCCACACTCATTTTTTAATATTT 1116
Qy 945 agttccagatctgtactgtgaccttctcacactgtagataaacattactcttttgc 1004
Db 1115 AGTTCACAGATCTGACTGTGACCTTTCTACACTGTAGAAATACATTTACTCATTTTGTTC 1056
Qy 1005 aaagaccctctgttgcctgaataatagctgactgttttccaaaggaggtgtctg 1064
Db 1055 AAAGACCCTTCTGTTGCTGCCTAATATCTAGCTGACTGTTTTTCTTAAGGAGTGTCTG 996
Qy 1065 gccaggggatctgtgaacaggtgggaagcatctcaagatcttccagggttatactta 1124
Db 995 GCCCAGGGGATCTGTGACAGGCTGGGAAGCATCTCAAGATCTTTCCAGGGTTATACTTGA 936
Qy 1125 ctgacacacagatgcatactacagagtgaaattatctaatcaacatccctcaggtct 1184
Db 935 CTAGCACACAGGATGATCATTTACGGAGTGAAATATCTAATCAACATCATCTCCAGTGTCT 876
Qy 1185 ttgccataactgaaattcatttccacatttgcaccttctcaagacctcaaaatgtca 1244
Db 875 TTGCCATACTGAATTTCAATTTCCCACTTTTGTGCCCACTTCTCAAGACCTCAAAATGTCA 816
Qy 1245 ttccatataatacacaggattaaacttttttttaacctgggaagaaattcaattgtacatg 1304
Db 815 TTCCATTAATATCACAGGATTAAC-TTTTTTTTAACTTGAAGAAATTCATGTTTACATG 757
Qy 1305 cagcatgggaatttaattacataattttttccagtgcaaaagatgactaagtccttta 1364
Db 756 CAGCATTTGGGAATTTAAATTTACATATTTTGTTTTCCAGTGCAAAAGATGACTAAGTCCCTTGA 697
Qy 1365 tccctccctcttgttgattttttttccagtgataaaagttaaaatgcttagcctgtgactg 1424

Db 696 TCCTCCCTTTGTTGATTTTTTTCAGTATATAAGTTAAAAATGCTTAGCCTTGACTG 637
QY 1425 agcgtgtataag-cacagcctctcccatccctccagccttatctgtcatcaccatcaa 1483
Db 636 AGCTGTATACAGCCACAGCCTCTCCCATCCCTCCAGCCTTATCTGTGTCATCACCATCAA 577
QY 1484 ccctcccataccacctaacaataacttgtaattccttgaacatgtcaggacata 1543
Db 576 CCCCTCCCAT-GCACCTAAACAAATCTAACTTGTATTCCTTGAACATGTCAGG-CATA 519
QY 1544 cattattcctctgctgagaagctcttctgtctcttaatactcgaatgatgtaagt 1603
Db 518 CATTATTCCTCTGCTGGAAGCTTCTCTCTCTCTTAAATCTAGATGATGTAAGT 459
QY 1604 ttgtaataagttgactatcttactcatcacaagaagagacacatgatgattccatc 1663
Db 458 TTTGAATAGTTGACATCTTCTTCATGCAAGAAGAGGACATATGATGATTCATCATC 399
QY 1664 acatgagacagcaataactaaaagtgttaattgattataagagtttagataaataatga 1723
Db 398 ACATGAGACAGCAAAATACATAAAGTGT-ATTTCATTATAAGAGTTTAGATAAATATGA 340
QY 1724 aatgcaagagcacaagaggaatgtttatggggcagctttgtaagcctgggatgtaagc 1783
Db 339 AATGCAAGAGCCACAGAGGGAATGTTTATGGGACGCTTGTAAAGCCTGGGATGTGAAGC 280
QY 1784 aaaggcagggaacctcatagttcttataataatacttcttctctctctatcaca 1843
Db 279 AAAGGCGGAACCTCATAGTATCTTATATATATATATCTTCACTTCTATCTATCACA 220
QY 1844 atatccacaagcttttcacagaattcagagtcagtcgcaaatccccaaagtgtaacctttatc 1903
Db 219 ATATCCAACAAGCTTTTTCAGAAATTCATGCAATTCATGCAATTCGCCAAAGTAACTTTATC 160
QY 1904 catttcagtgagtcgcttttagaatttttgccaatactactggtcacttatctcaact 1963
Db 159 CATTTCATGGTGAGTCGCTTTTGAATTTTGGCAAAATCATCTGTCACCTTATCTCAACT 100
QY 1964 ttgagatgtgttctctgttagttaattgaaagaaatagggcaactcttctgtgagccact 2023
Db 99 TTGAGATGTGTTGCTCTGTAGTTAAATGAAAGAAATAGGGCACTCTTGTGAGCCACTT 40
QY 2024 tagggttcaactcctggcaataaagaatttacaagaagct 2062
Db 39 TAGGGTTCACTCTCTGGCAATAAAGAAATTTACAAAGACT 1

RESULT 7

AAV62429
ID AAV62429 standard; cdna; 820 BP.
XX AC AAV62429;
XX XX
XX 30-DEC-1998 (first entry)
DE DE
DE Prostate cancer antigen (PCA3) cdna splice variant 3.
KW Prostate cancer antigen cdna splice variant 3; PCA3; prostatic cancer;
KW PC; ds.
XX
OS Homo sapiens.
XX
XX WO9845420-A1.
XX
PD 15-OCT-1998.
XX
XX 09-APR-1998; 98WO-CA00346.
XX
XX 10-APR-1997; 97US-0041836.
XX
PA (DIAG-) DIAGNOCURE INC.

PI Bussemakers MJG;
XX WPI: 1998-568347/48.
XX
PT New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
PT prevention and treatment of prostatic cancer
XX
PS Claim 4: Pages 77-78; 111pp; English.
XX
CC The present sequence represents the prostate cancer antigen (PCA3)
CC cdna splice variant 3 sequence comprising of exons 1, 3, and 4a
CC of the PCA3 gene. The PCA3 cdna splice variant 3 sequence,
CC isolated from a human primary prostatic tumour tissue cdna library,
CC was found in approximately 15% of the cdna clones isolated. The
CC invention claims for PCA3 cdna variants and the proteins they encode.
CC The invention also claims for antibodies against PCA3 protein. The
CC antibodies are claimed to be useful for detecting PCA3 protein in
CC immunosay tests, for diagnosing, assessing and prognosing of
CC prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin
CC or radioisotope, and nucleic acids antisense to PCA3 cdna are claimed
CC to be useful for treating PC, while determining elevated levels of
CC PCA3 (as RNA or protein) is useful for detecting a predisposition
CC to development of PC, e.g. in prenatal tests. Detecting PCA3 protein
CC allows differentiation between malignant and benign prostatic disease,
CC and the level of PCA3 expression allows correlation with the grade of
CC tumour. PCA3 protein and its fragments are also claimed to be useful
CC in vaccines for preventing PC; in drug screens for identifying
CC specific (ant)agonists (potentially useful therapeutically) and for
CC studying protein-DNA interactions.
XX
SQ Sequence 820 BP; 262 A; 169 C; 191 G; 198 T; 0 other;

Query Match 20.3%; Score 725.4; DB 19; Length 820;
Best Local Similarity 99.9%; Pred. No. 4.3e-189;
Matches 726; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 281 caggggtgagaataaagaaggctgctgactttaccatcttgaggccacacatctgtgaa 340
Db 94 cagaggtgagaataaagaaggctgctgactttaccatcttgaggccacacatctgtgaa 153
QY 341 atggagataattaaactacactagaacacagacagatgacaataatgtctaagtagtac 400
Db 154 atggagataattaaactacactagaacacagacagatgacaataatgtctaagtagtac 213
QY 401 atgtttttgcacatttccagccctttaaataatcacacacagagaaagagaa 460
Db 214 atgtttttgcacatttccagccctttaaataatcacacacagagaaagagaa 273
QY 461 gcacagagatccctgggagaataatgccggcccatcttgggtcatcgatgagcctcgcc 520
Db 274 gcacagagatccctgggagaataatgccggcccatcttgggtcatcgatgagcctcgcc 333
QY 521 ctgtgctgtgtccctgtgagggagacattagaaaaatgaatgtgtgttccttaa 580
Db 334 ctgtgctgtgtccctgtgagggagacattagaaaaatgaatgtgtgttccttaa 393
QY 581 aggatgggcagaaacacagatcctgtgtgatatattttgaacgggattacagatttg 640
Db 394 aggatgggcagaaacacagatcctgtgtgatatattttgaacgggattacagatttg 453
QY 641 aaatgaagtcaaaaagtgcattaccatgagaggaacacagacagaaaaatcttgatg 700
Db 454 aaatgaagtcaaaaagtgcattaccatgagaggaacacagacagaaaaatcttgatg 513
QY 701 gttcacagaacatgcacaacaaaaatggaatactgtgatgacatgagggagcagaagt 760
Db 514 gttcacagaacatgcacaacaaaaatggaatactgtgatgacatgagggagcagaagt 573
QY 761 ggggagagataaacacaggggcagaggggtccaggatttctgacctgctctaaactgtgc 820
Db 574 ggggagagataaacacaggggcagaggggtccaggatttctgacctgctctaaactgtgc 633

QY 821 gttcataacaaatcatttcatttcataatttcataacccctcaaaacaaagctgtgtgtaatactga 880
Db 634 gttcataacaaatcatttcatttcataatttcataacccctcaaaacaaagctgtgtgtaatactga 693
QY 881 tctctacgggttcctctctggtggcccaacattctccatatatccagccacacactcaattttta 940
Db 694 tctctacgggttcctctctggtggcccaacattctccatatatccagccacacactcaattttta 753
QY 941 atttagttccagatctgactgtgacctttctacactgtagaataaacattactcatttt 1000
Db 754 atttagttccagatctgactgtgacctttctacactgtagaataaacattactcatttt 813
QY 1001 gttcaaa 1007
Db 814 gttcaaa 820

RESULT 8
ID AAA06690/c
XX AAA06690 standard; cDNA; 812 BP.
AC AAA06690;
XX
DT 13-JUN-2000 (first entry)
XX
DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:471.
XX
KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW immunogenic; cytostatic; vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200004149-A2.
XX
PD 27-JAN-2000.
XX
PF 14-JUL-1999; 99WO-US15838.
XX
PR 14-JUL-1998; 98US-0115453.
PR 14-JUL-1998; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX
PA (CORI-) CORIXA CORP.
XX
PI Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;
XX
WPI; 2000-171268/15.
XX
PT New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumor protein -
XX
PS
PS Claim 1; Page 262; 263pp; English.

XX
CC The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate antibodies
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the polynucleotides encoding the polypeptides can be used as a probe or
CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC AAA82000 to AAA82020 represent sequences used in the exemplification of
CC the present invention.

XX
SQ Sequence 812 BP; 198 A; 189 C; 169 G; 256 T; 0 other;

Query Match 20.2%; Score 722.4; DB 21; Length 812;
Best Local Similarity 99.9%; Pred. No. 2.8e-188;
Matches 723; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 281 caggggtgagaaataagaagagctgctgactttacacatctgaggccacacatctcctgaa 340
Db 724 CAGAGGTGAGAAATAGAAAGGCTGCTGACTTTTACCATCTGAGGCCACACATCTCTGAA 665
QY 341 atggagataataacatcactagaacagcaagatgacaataataatgtctaaagtacac 400
Db 664 ATGGAGATAAATTAACATCACTAGAACAGCAAGATGACAATAATATGCTTAAGTAGTGAC 605
QY 401 atgttttgcacattccagcccttttaatatccacacacaggaagacacaaaggaa 460
Db 604 ATGTTTTTGCACATTTCCAGCCCCCTTTAAATATCCACACACACAGGAAGCACAAGGAA 545
QY 461 gcacagagatccctgggagaaatgcccgccgcctcttgggtcatcgatgagcctcgcc 520
Db 544 GCACAGAGATCCCTGGGAGAAATGCCCGCCGCCATCTTGGGTCTATCGATGAGCCTCGCC 485
QY 521 ctgtgcctggtcccgctgtgaggaagacattagaataatgaattgattgttctctaa 580
Db 484 CTGTGCTGGTCCCGCTGTGAGGGAAGGACATTAGAAATGAATGATGTGTCTCTTAA 425
QY 581 aggtgggcaggaacacagatcctgtgtggatattttgaacgggattacagatttg 640
Db 424 AGGATGGCGCAGGAAACAGATCCTGTTGGGATATTATTTCACCGGGATTACAGATTG 365
QY 641 aaatgaagtacaaagtgcattaccatgagaggaacacagacagaaaaatcttgatg 700
Db 364 AAATGAAGTCACAAAGTGAGCATTTACCAATGAGAGGAAACAGACGAGAAAAATCTTGATG 305
QY 701 gttcacagacatgcaacaaacaaatgaatactgtgatacatgacatgagcgcaagct 760
Db 304 GCTTCACAGACATCAACAAACAAATGGAATCTGTGATGACATGAGGCGAGCCAAAGCT 245
QY 761 ggggaggagataaccacgggagaggtcaggattctgtgcccctgctgctaaactgtgc 820
Db 244 GGGGAGGAGATAACCCAGGGGAGAGGTCAGGATCTGCCCCCTGCTGCTAACTGTGC 185
QY 821 gttcataacaaatcatttcatttcataacccctcaaaacaaagctgtgtgtaatactga 880
Db 184 GTTCATAACCAATCATTTTCATATTCTAACCTCAAAACAAAGCTGTTGTAATCTGCA 125
QY 881 tctctacgggttcctctctggtggcccaacattctccatatatccagccacacactcatttt 940
Db 124 TCTCTACGGTTCCTTCTGGGCCCAACATTTCTCATATATCCAGCCACACTCAATTTTAA 65
QY 941 atttagttccagatctgactgtgacctttctacactgtagaataaacattactcatttt 1000
Db 64 ATTTAGTTCCAGATCTGACTGTGACCTTTCTACACTGTAGAAATACATTACTCATTTT 5
QY 1001 gttc 1004
Db 4 GTTC 1

RESULT 9
ID AAX37486
XX AAX37486 standard; cDNA; 597 BP.
AC AAX37486;
XX
DT 06-JUL-1999 (first entry)
XX
DE Human secreted protein cDNA fragment containing gene 36.
XX
KW Human; secreted protein; treatment; prevention; protein therapy; AIDS;
KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW immune system disease; autoimmune disease; hepatic disease; lymphoma;
KW renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;

KW cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
KW pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
XX arthritis; malignancy; digestive; endocrine; infection; ss.

OS Homo sapiens.

PN W09918208-AL.

XX 15-APR-1999.

XX 01-OCT-1998; 98WO-US20775.

XX 02-OCT-1997; 97US-0060884.

XX 02-OCT-1997; 97US-0060833.

PR 02-OCT-1997; 97US-0060836.

PR 02-OCT-1997; 97US-0060837.

PR 02-OCT-1997; 97US-0060838.

PR 02-OCT-1997; 97US-0060839.

PR 02-OCT-1997; 97US-0060843.

PR 02-OCT-1997; 97US-0060862.

PR 02-OCT-1997; 97US-0060866.

PR 02-OCT-1997; 97US-0060874.

XX (HUMA-) HUMAN GENOME SCI INC.

PA Carter KC, Duan DR, Endress GA, Feng P, Ferrie AM;

XX Florence KA, Greene JM, Janat F, Lafleur DW, Ni J;

PI Rosen CA, Ruben SM, Shi Y, Young P, Yu G;

XX WPI: 1999-264022/22.

DR P-PSDB; AAY07887.

XX New isolated human genes and the secreted polypeptides they encode

XX Claim 1a; Page 247; 368pp; English.

XX This invention describes novel isolated human genes and the secreted

CC proteins they encode. The products of the invention are useful for

CC preventing, treating or ameliorating medical conditions, e.g. by protein

CC or gene therapy. Also pathological conditions can be diagnosed by

CC determining the amount of the new polypeptides in a sample or by

CC determining the presence of mutations in the new polynucleotides.

CC Specific uses are described for each of the 101 polynucleotides, based

CC on which tissues they are most highly expressed in, and include

CC developing products for the diagnosis or treatment of cancer, tumours,

CC neurodegenerative disorders, developmental abnormalities and fetal

CC deficiencies, blood disorders, leukemias, diseases of the immune system,

CC autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,

CC allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate

CC disease, skeletal or cardiac muscle disorders, pulmonary disorders,

CC transplant rejection, disorders involving osteoclasts such as

CC osteoporosis, arthritis or malignancies, digestive/endocrine disorders,

CC infections and AIDS. The human secreted proteins of the invention are

CC represented in AAY07852-Y07993 and the encoding nucleic acids are

XX represented in AAX37451-X37552.

XX Sequence 597 BP; 181 A; 131 C; 134 G; 150 T; 1 other;

Query Match 14.6%; Score 524; DB 20; Length 597;
Best Local Similarity 99.6%; Pred. No. 7.3e-134;
Matches 524; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 484 gcccgccgcacattgggtcattcgatgagcctcgccctgtgctgtgctccgctgttgag 543

|||||

Db 54 gcccgccgcacattgggtcattcgatgagcctcgccctgtgctgtgctccgctgttgag 113

|||||

QY 544 ggaagacattagaaatgaattgattgttctcttaagagtgaggcagaaacagatcc 603

|||||

Db 114 ggaagacattagaaatgaattgattgttctcttaagagtgaggcagaaacagatcc 173

|||||

QY 604 tgttggatatttttaacgggattcacagattggaattgaagtcacaaagtggagcat 663

|||||

Db 174 tgttggatatttttaagcgggttacagatttgaatgaagtcacaaagtggagcat 233

QY 664 taccatgagagaaacacagacagagaaatcttgatggtctcacaagacatgcacaaac 723

|||||

Db 234 taccatgagagaaacacagacagagaaatcttgatggtctcacaagacatgcacaaac 293

|||||

QY 724 aaaaatgaatactgtgatgacatgagcgagcgaagctggggagagataaccacggggca 783

|||||

Db 294 aaaaatgaatactgtgatgacatgagcgagcgaagctggggagagataaccacggggca 353

|||||

QY 784 gagggtcaggattctgcccctgctgctaactgacggttcataaccacaaatcttcata 843

|||||

Db 354 gagggtcaggattctgcccctgctgctaactgacggttcataaccacaaatcttcata 413

|||||

QY 844 ttctaacctcaaaaacagctgttgtaatactgacgtctacgttctctctgagccc 903

|||||

Db 414 ttctaacctcaaaaacagctgttgtaatactgacgtctacgttctctctgagccc 473

|||||

QY 904 aacattctccatatatccagccacactcatttttaataatttagttccagatctgtactg 963

|||||

Db 474 aacattctccatatatccagccacactcatttttaataatttagttccagatctgtactg 533

|||||

QY 964 tgacctttctacactgtagaataacattactcattttgttcaaaa 1009

|||||

Db 534 tgacctttctacactgtagaataacattactcattttgttcaaaa 579

|||||

RESULT 10
AAA06545
ID AAA06545 standard; cDNA; 718 BP.
XX
AC AAA06545;
XX
DT 13-JUN-2000 (first entry)
XX
DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:313.
XX
KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW immunogenic; cytostatic; vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200004149-A2.
XX
DT 27-JAN-2000.
XX
PF 14-JUL-1999; 99WO-US15838.
XX
PR 14-JUL-1999; 98US-0115453.
PR 14-JUL-1999; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX
PA (CORI-) CORIXA CORP.
XX
PI Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;
XX
DR WPI: 2000-171268/15.
XX
PT New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumor protein -
XX
PS Claim 1; Page 199-200; 263pp; English.
XX

CC The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines


```
XX Homo sapiens.
OS
PN WO200004149-A2.
XX
PD 27-JAN-2000.
XX
PF 14-JUL-1999; 99WO-US15838.
XX
PR 14-JUL-1998; 98US-0115453.
PR 14-JUL-1998; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX
PA (CORI-) CORIXA CORP.
XX
PI Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;
XX
DR WPI; 2000-171268/15.
XX
XX New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumor protein -
XX
PS Claim 1; Page 192; 263pp; English.
XX
XX The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate antibodies
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the polynucleotides encoding the polypeptides can be used as a probe or
CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC AAY82000 to AAY82020 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 301 BP; 76 A; 58 C; 70 G; 97 T; 0 other;

Query Match 8.1%; Score 288.4; DB 21; Length 301;
Best Local Similarity 99.3%; Pred. No. 2.5e-69;
Matches 300; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 659 agcattaccaatgagaggaacacagacgagaaaattcttgatggcttcacaagacatgcaa 718
DB 301 AGCATTACCAATGAGAGGAAACACAGACGAGAAAATCTTGATGCTTCACAAAGACATGCAA 242
QY 719 caaacaaaatggaatactgtgatgacatgagcagcagcaagctgggagagataaacacg 778
DB 241 CAACAAAATGGAATCTGTGATAACATGAGGAGCCCAAGCTGGGAGGAGATAAACACG 182
QY 779 ggcagagggctcaggattctgcctcgtcgtcctaaactgtgcttcataaccacaaatcatt 838
DB 181 GGCAGAGGGTCAAGGATCTGGCCCTGCTGCTTAAACTGTGCTTTCATAACCAATCAAT 122
QY 839 tcattattctaacccctcaaaacaaaagctgtgtgtaatatctgatctcaggttcctcttg 898
DB 121 TCATATTCTTAACCCCTCAAAACAAAGCTGTGTGTAATATCTGATCTCTAC-GTTCCCTCTG 63
QY 899 ggcacaaattctccatatatccagccacaaactcatttttaattagttccagatctg 958
DB 62 GGCCCAACATCTCCATATATCCAGCCACACTCATTTTAAATATTAGTTCACAGATCTG 3
QY 959 ta 960
DB 2 TA 1
```

```
RESULT 13
AAA06468
ID AAA06468 standard; cDNA; 283 BP.
XX
AC AAA06468;
XX
DT 13-JUN-2000 (first entry)
XX
XX Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:235.
XX
KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
XX immunogenic; cytostatic; vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200004149-A2.
XX
PD 27-JAN-2000.
XX
PF 14-JUL-1999; 99WO-US15838.
XX
PR 14-JUL-1998; 98US-0115453.
PR 14-JUL-1998; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX
PA (CORI-) CORIXA CORP.
XX
PI Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;
XX
DR WPI; 2000-171268/15.
XX
XX New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumor protein -
XX
PS Claim 1; Page 179; 263pp; English.
XX
XX The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate antibodies
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the polynucleotides encoding the polypeptides can be used as a probe or
CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC AAY82000 to AAY82020 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 283 BP; 84 A; 50 C; 60 G; 89 T; 0 other;

Query Match 6.9%; Score 247; DB 21; Length 283;
Best Local Similarity 98.6%; Pred. No. 5.7e-58;
Matches 281; Conservative 0; Mismatches 0; Indels 4; Gaps 3;

QY 2425 tggggcgtgcatcaggc-ggtttgagaaatattcattctcagcagaagccagaatttg 2483
DB 1 tggggcgtgcatcaggcgggtttgagaaatattcattctcagcagaagccagaatttg 60
QY 2484 aattccctcatctttka-ggaatcattaccaggtttggagaggattccagacagctcag 2542
DB 61 aattccctcatcttttaggggaatcattaccaggtttggagaggattccagacagctcag 120
QY 2543 tgctttcaataatgtctcgaactctgcctcttctgttctcagatgagtcataaa 2602
DB 121 tgctttcaataatgtctcgaactctgcctcttctgttctcagatgagtcataaa 178
```



```
Qy 2603 taattattcttgaactgctcataggagagagataataagaactctgagtgatatcaa 2662
|||||
Db 179 taattattcttgaactgctcataggagagagataataagaactctgagtgatatcaa 238
|||||
Qy 2663 cattaggattcaagaataattagattgaagctcacactggtca 2707
|||||
Db 239 cattaggattcaagaataattagattgaagctcacactggtca 283
|||||

RESULT 14
AAZ33445
ID AAZ33445 standard; cDNA; 359 BP.
XX
AC AAZ33445;
XX
DT 08-DEC-1999 (first entry)
XX
DE Human prostate cancer-associated EST 23.
XX
KW Expressed sequence tag; EST; prostate tumor; antitumor; treatment;
KW gene therapy; tissue specificity human; ss.
XX
OS Homo sapiens.
XX
PN DE19811193-A1.
XX
PD 16-SEP-1999.
XX
PF 10-MAR-1998; 98DE-1011193.
XX
PR 10-MAR-1998; 98DE-1011193.
XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;
XX
DR WPI; 1999-519628/44.
XX
DR P-PSDB; AAY48243.
XX
PT New nucleic acid expressed at high level in prostatic tumor tissue and
PT encoded polypeptides, useful for treating cancer and screening for
PT therapeutic agents.
XX
PS Claim 1a; 87; 166pp; German.
XX
CC This invention describes novel nucleic acid sequences (A) that are
CC expressed at high level in prostatic tumor tissue and encode gene
CC products or their fragments. The products of the invention have
CC antitumor activity. Polypeptides (I) encoded by (A) are used: (i) for
CC identifying agents for treatment of prostatic cancer and (ii) for
CC therapy of prostate cancer, optionally where expressed by gene therapy
CC methods. (A) is also used to isolate full-length genes (for gene therapy)
CC and for recombinant production of (I), which can be used to raise
CC specific antibodies. (A) are identified by assembly of ESTs (expressed
CC sequence tags) before they are analyzed for expression pattern (tissue
CC specificity). This approach eliminates many of the false results, as
CC regards tissue specificity, associated with known methods that use
CC single (usually short) ESTs. AAZ33423-233476 represent expressed
CC sequence tags described in the method of the invention.
XX
SQ Sequence 359 BP; 121 A; 75 C; 94 G; 69 T; 0 other;

Query Match 6.6%; Score 235.4; DB 20; Length 359;
Best Local Similarity 99.6%; Pred. NO. 1e-54;
Matches 236; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 281 cagggggtgagaaataagaaggctgctgactttaccatctgagggccacacatctgctgaa 340
|||||
Db 123 cagaggtgagaaataagaaggctgctgactttaccatctgagggccacacatctgctgaa 182
|||||
Qy 341 atgggagataataacatcactagaaacagcaagatgacaataataatgtctaagtgtgac 400
|||||
```

```
Db 183 atgggagataataacatcactagaaacagcaagatgacaataataatgtctaagtgtgac 242
|||||
Qy 401 atgtttttgcacatttccagcccttttaataatccacacacacaggaagcacaagaaggaa 460
|||||
Db 243 atgtttttgcacatttccagcccttttaataatccacacacacaggaagcacaagaaggaa 302
|||||
Qy 461 gcacagagatccctgggagaaatgcccgccgccatcttgggtcatcgatgagcttc 517
|||||
Db 303 gcacagagatccctgggagaaatgcccgccgccatcttgggtcatcgatgagcttc 359
|||||

RESULT 15
AAF58252/c
ID AAF58252 standard; DNA; 936 BP.
XX
AC AAF58252;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1835.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface.
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 3.7%; Score 133; DB 22; Length 936;
Best Local Similarity 0.9%; Pred. NO. 2.3e-26;
Matches 7; Conservative 490; Mismatches 280; Indels 0; Gaps 0;

Qy 1154 aattatctaataacatcctcagtgctttggccatactgaaattcatttccactt 1213
|||||
Db 777 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 718
|||||
Qy 1214 ttgtgccattctcaagacctcaaaatgtcattccatttaataatcacagattaaactttt 1273
|||||
Db 717 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 658
|||||
Qy 1274 ttttaacctggaagaattcaatgttcatgcagctatgggaatttaatacatatttg 1333
|||||
Db 657 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 598
|||||
```

Search completed: July 31, 2001, 07:27:42
Job time: 17874 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2001, 08:15:33 ; Search time 10902.7 Seconds
(without alignments)
5081.817 Million cell updates/sec

Title: US-09-402-713A-6
Perfect score: 3582
Sequence: 1 acagaagaataagcaagtgc.....tgattctttgttacaacttt 3582

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 1344157 seqs, 7733874588 residues 2688314
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vil:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3497	97.6	3923	88	AF103907	AF103907 Homo sapi
2	3032.4	84.7	5435	88	AF103908	AF103908 Homo sapi
3	2959.8	82.6	164371	80	AL390239	AL390239 Homo sapi
4	1363.4	38.1	173831	80	AL359314	AL359314 Homo sapi
c 5	1319.8	36.8	267581	80	AL358573	AL358573 Homo sapi
c 6	536	15.0	546	54	G56925	G56925 SHGC-102485
c 7	533	14.9	542	54	G56926	G56926 SHGC-102486
8	235.4	6.6	359	9	AX018075	AX018075 Sequence

```
9 121 3.4 143675 79 AL161625 Homo sapi
10 121 3.4 267581 80 AL358573 Homo sapi
11 120 3.4 580 89 AF279290 Homo sapi
12 119.4 3.3 172298 61 AC009556 Homo sapi
13 77.4 2.2 7218 10 I66494 Sequence 14
14 57.4 1.6 7218 10 I66494 Sequence 14
15 56 1.6 163266 70 AC027016
16 53.8 1.5 502 7 BTMISATF
17 53 1.5 1141 10 AX083744
18 52.2 1.5 86064 81 AL391554 Homo sapi
19 52.2 1.5 143489 79 AL161637 Homo sapi
20 51.6 1.4 182999 63 AC013260 Homo sapi
21 50.8 1.4 161222 64 AC016037 Homo sapi
22 50.8 1.4 169942 80 AL365230 Homo sapi
23 50.4 1.4 118981 71 AC040983 Mus muscu
24 50.2 1.4 153054 81 AL450423 Homo sapi
25 50 1.4 200000 60 AC007705 Homo sapi
26 49.2 1.4 199563 82 AL590503 Mus muscu
27 48.8 1.4 291 53 AU025050 Rattus no
28 48.8 1.4 201885 66 AC020703 Homo sapi
29 48.6 1.4 112902 85 AC005230 Homo sapi
30 48.6 1.4 145554 84 CNS07E0U Homo sapi
31 48.4 1.4 155103 76 AC083837 Homo sapi
32 48.4 1.4 167075 92 HS742324 Human DNA
33 48.4 1.4 186572 75 AC079130 Mus muscu
34 48.2 1.3 113367 86 AC008178 Homo sapi
35 48 1.3 157676 70 AC027358 Homo sapi
36 48 1.3 180392 65 AC019299 Homo sapi
37 48 1.3 205053 87 AC011749 Homo sapi
38 47.8 1.3 68863 71 AC036112 Homo sapi
39 47.4 1.3 74266 77 AC087104 Arabidops
40 47.4 1.3 80376 12 AB010073 Arabidops
41 47.4 1.3 130705 92 HS232122 Homo sapi
42 47.2 1.3 1141 10 AX083744 Sequence
43 47.2 1.3 137064 74 AC073363 Homo sapi
44 47.2 1.3 152571 66 AC021589 Homo sapi
45 47.2 1.3 160133 69 AC025140 Homo sapi
```

ALIGNMENTS

```
RESULT 1
AF103907 3923 bp mRNA PRI 14-AUG-2000
LOCUS Homo sapiens non-coding RNA DD3 sequence.
ACCESSION AF103907
VERSION AF103907.1 GI:6165973
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3923)
Bussemakers,M.J., van Bokhoven,A., Verhaegh,G.W., Smit,F.P.,
Karthaas,H.F.M., Schalken,J.A., Debruyne,F.M., Ru,N. and Isaacs,W.B.
DD3: a new prostate-specific gene, highly overexpressed in prostate
cancer
Cancer Res. 59 (23), 5975-5979 (1999)
JOURNAL 20072260
MEDLINE 10606244
PUBMED
2 (bases 1 to 3923)
Bussemakers,M.J.G., Van Bokhoven,A., Verhaegh,G.W., Smit,F.P.,
Karthaas,H.F.M., Schalken,J.A., Debruyne,F.M.J., Ru,N. and
Isaacs,W.B.
Direct Submission
Submitted (28-OCT-1998) Urology Research Laboratory, University
Hospital Nijmegen, P.O. Box 9101, Nijmegen 6500 HB, Netherlands
Location/Qualifiers
1..3923
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
```

|||||
Db 841 ATATTCTAACCCCTAAAACAAGCTGTTGTAATATCTGATCTACGGTTCCTCTGGG 900
QY 901 cccaattctccatataccaagccacactcattttaaatttagttccceagatctgta 960
Db 901 CCCAACATTCTCCATATATCCAGCCACACTCATATTTAATAATTTAGTTCCTCCAGATCTGTA 960
QY 961 ctgtgacctttctacactgtagaataaacattactcatttctgttccaagacctctggtt 1020
Db 961 CTGTGACCTTTCTACACTGTAGATAAACATTACTCATTTGTTCAAAGACCCCTCGTGT 1020
QY 1021 gctgctaataatgtagctgactgttttctctaagagtgcttctgcccaggggactgtg 1080
Db 1021 GCTGCCATAATATGTAGCTGACTGTTTTTCTCTAAGGAGTGTCTGGCCAGGGGATCTGTG 1080
QY 1081 aacaggctgggaagcatctcaagatctttccagggttatacttactagcacacagcatga 1140
Db 1081 AACAGGCTGGGAAGCATCTCAAGATCTTTTCCAGGGTTATACTTAGCACACAGCATGA 1140
QY 1141 tcattacggagtgaattatcttaactcaacatcatctcagtgctctttggccatactgaaat 1200
Db 1141 TCATTACGGAGTGAATTATCTAATCAACATCATCTCAGTGTCTTTGGCCCATACTGAAT 1200
QY 1201 tcatttcccactttgtgcccattctcaagacctcaaaaatgtcattccattaatatcaca 1260
Db 1201 TCATTTCCCACTTTTGTGCCATTTCTCAAGACCTCAAAATGTCATTCCCATTAATATACA 1260
QY 1261 gattataacttttttttaacctggaagaattcaattgaatttgaatgcatgagctatgggaattta 1320
Db 1261 GGATTAACTTTTTTTTTAACTGGGAATTTCAATGTTTACATGACAGCTATGGGAATTTA 1320
QY 1321 attacataatttgtttccagtgcaagatgactcctcagtgctcttaccctcccctttgttt 1380
Db 1321 ATTACATATTTTTTCAGTGCAGGAAGATGACTAAGTCTCTTATCCCTCCCTTTGTTT 1380
QY 1381 gattttttccagtaataagttaaaatgcttagccttgtagccttgtagggctgtatagacac 1440
Db 1381 GATTTTTTTTCCAGTATAAAGTTAAATGCTTAGCCTTACTAGGCTGTATACAGCAC 1440
QY 1441 agcctctcccactccctcagcttatctgtcatcaccatcaacccctcccataccact 1500
Db 1441 AGCCTCCTCCATCCCTCCAGCTTATCTGTATCACCATCAACCCCTCCCATACCACCT 1500
QY 1501 aacaaaaactaacttgtaattccttgaacatgtccaggaacatacatattcctctgct 1560
Db 1501 AAACAAAATCTAATTTGTAATCTTGAACATGTCAGGACATACATTATTCTCTGCT 1560
QY 1561 gagaagcttctctgtctcttaaatcttagaatgtataagttttgataagttgacta 1620
Db 1561 GAGAAGCTCTCTCTGCTCTTTAAATCTAGAATGATGTAAGTTTGTAAAGTTTGACTTA 1620
QY 1621 tcttacttcatgaagaaggacacatatgagattcatcatcactgagacagcaata 1680
Db 1621 TCTTACTTCATGCAAGAAGGACACATATGAGTTTCATCATCATGAGACAGCAATA 1680
QY 1681 ctaaaagtgaatttgaattagaagtttagataaataatgaaatgcaagaccacaga 1740
Db 1681 CTAAAAAGTGAATTTGATTTAAGAGTTTAGATAAATATGNAATGCAAGGCCACAGA 1740
QY 1741 gggaaatgtttatggggcactgttgaagcctgggagtgtgaagcaaaaggcaggaaacctca 1800
Db 1741 GGGAAATGTTTATGGGGCAGCTTGTAAAGCTGGGATGTGAAGCAAAAGGCGAGCACTCA 1800
QY 1801 tagtatcttataataataactcatttctctatctctatcaataatccaaagctttt 1860
Db 1801 TAGTATCTTATATAATATCTTCTATTTCTATCTATCAATAATCAACAAGCTTTT 1860
QY 1861 cacagaaticatgcagtgcaaatccccaaaggtaacctttatccatttcatgtgagtgagtc 1920
Db 1861 CACAGAAATTCATGCAGTGCANAATCCCCAAAGGTAACTTTATCCATTTTCATGGTAGTGC 1920
QY 1921 gctttagaattttggcaaatcatactggtgcaacttatctcaactttgagatgtgtttgtcc 1980
|||||

Db 1921 GCTTTAGAAATTTTGGCAATCATACTGTGCTACTTATCTCAACTTTGAGATGTGTTGTCC 1980
QY 1981 ttgtagttaattgaagaataagcactctgtgaagcaacttttaggtttcactcctggc 2040
Db 1981 TTGTAGTTAATTTGAAGAAATAGGGCACHCTTGTGAGCCACTTTTAGGGTTCACCTCTGGC 2040
QY 2041 aataagaattttacaagaagctactcagaccagttgtttaagagctctgtgtgtgtgt 2100
Db 2041 AATAAGAATTTTACAAAGAGCTACTCAGGACCAGTTGTTAAGAGCTCTGTGTGTGTGT 2100
QY 2101 ggtgtgtgtgagttacatgccaagaagtgtgctctctctctctgtaccatatttccagac 2160
Db 2101 GT 2160
QY 2161 ttaaaacagcatgttttcaaatgacactatgctgccaatgactgtgccaatgactgtatccacacatat 2220
Db 2161 TTAAACAAGCATGTTTTCAATGGCACHTATGAGCTGCCCAGTATGATACACACCATAT 2220
QY 2221 ctcatattctccagtaaatgtgataaattgtcatctgtttaacataaaaaaagtttgc 2280
Db 2221 CTCATTATTCTCCAGTAAATGTGATAAATATGTCATCTCTTAACATAAAAAAAGTTTGAC 2280
QY 2281 ttcaaaaagcagctggaaaatggacaaccaaatatgcaataatctaaactcctaccatca 2340
Db 2281 TTCAAAAAAGCAGCTGGAATGGCAACCAACCAATATGCAATAAATCTTAACCTCTACCATCA 2340
QY 2341 gctcacactgctgacatatattgttagaagcacctcagattgtggttctctcttaagc 2400
Db 2341 GCTACACACTGCTTGACATATATTTTGAAGCACCCTCGATTTGGGGTCTCTCTTAAGC 2400
QY 2401 aaaaacttctcagttcagctcagctgggctgtgcatcagcaggcgttttgagaaatttcaa 2460
Db 2401 AAAATACTTTGCATTTAGTCTCAGCTGGGCTGTGTCATCAGCGGTTTGAAGAAATATTCAA 2460
QY 2461 ttctcagaagaagcagaatttgaattccctcactctctttaggaataatttaccaggttg 2520
Db 2461 TTCTCAGCAGAAAGCCAGAAATTTGAATTCCTCATCTCTTTTAGGAATCATTTACCAGGTTG 2520
QY 2521 gagagattcagacagctcaggtgctttcactaatgtctctgaaactctgtcctctctt 2580
Db 2521 GAGAGATTCAGACAGCTCAGGTGCTTTTCACTAATGTCTCTGNACTTCTGCTCTTTG 2580
QY 2581 tgttcagttagtcccaataaataatgttattctttgaactgatgctcatagagagaata 2640
Db 2581 TGTTCATGTGATAGTCCAAATAAATATGTTATCTTTGAAGTGTCTCATAGGAGAGATA 2640
QY 2641 taagaactctgagtgatatacaacattagggattccaagaataattagatttaagctccaca 2700
Db 2641 TAAGAACTCTGAGTGATATCAACATTTAGGGATTTCAAGAAATATTAGATTTAAGCTCACA 2700
QY 2701 ctggtcaaaaggaaccaagatacaagaactctgagctgctcatcgccccatctctgtga 2760
Db 2701 CTGGTCAAAAGGAACCAAGATACAAAGAACTGTGAGCTGTCTGTCGCCCATCTCTGTGA 2760
QY 2761 gccacaaccaacagcaggaccaccaacgcagctgtctgagatcctttaaatacaaggaaacagtg 2820
Db 2761 GCCACAACCAACAGCAGGACCACCAACGCTGTGTGAGATCTTTAAATCAAGGAACCAAGTG 2820
QY 2821 tcatgagttgaattctctctattatggatgctgctctgagctctgagcactctctgctctctct 2880
Db 2821 TCATGAGTTGAAATCT 2880
QY 2881 gacacatatagcttctagcctttgtctccagacttttctctctccacacacatcgc 2940
Db 2881 GACACATAATAGCTTCTAGCCTTTGCTTCCAGACTTTTATCTTTTCTCCACACATCGC 2940
QY 2941 ttaccaactctctctctgtctgtgtcttcttgacttcccccaagaatttccaagcactct 3000
Db 2941 TTACCAATCT 3000
QY 3001 caagtcttttctccatccccaccactaaacctgaattgctagacccttatttttattaa 3060
Db 3001 CAAGTCTTTTCTTCCATCCCACTAACTGAA-TGCCTAGACCCTTATTTTTATTAA 3059
|||||

```

QY 3061 ttccaatagatgctgcctatgggctaataattgcttttagatgaacattagattttaag 3120
Db 3060 TTTCCAATAGATGCTGCTATGGCT-ATATTGCTTTAGATGAACATTAGATATTTAAAG 3118
QY 3121 tctaaagaggttcaaaatccaactcaattattctcttcttcttcaactccctctcctct 3180
Db 3119 CTCAGAGGTTCAAAATCCAACATCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3177
QY 3181 cctatatactactatg-actgaacaggatgtcccaaa-gatgccagcaaatgagaaa 3238
Db 3178 CCTATATTAAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3237
QY 3239 ccagtggtcctcttggtgatcatgcatgaagactgctgaagccag-aggatgactgatt 3297
Db 3238 CCAGTGCTCTCTGTGGTACATGCATGCAAGACTGCTGAAGCCAGAGGATGACTGATT 3297
QY 3298 acgctcatggttgaggaggaccactcctcctcctcctcctcctcctcctcctcctcctcct 3357
Db 3298 ACGCCCTCATGGGTGGAGGGGACCACCTCTGGGCTTTCGTGATTGTCAGGAGCAAGACCTG 3357
QY 3358 agatgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 3417
Db 3358 AGATGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3417
QY 3418 ttgctacatttgagaattccaattaggaactcacatgctgttttctcctcctcctcctcctcctcct 3477
Db 3418 TTGCTACATTTGAGAAATCCAATTAGGAACCTCACATGTTTATCTGCCCTCAATTTT 3477
QY 3478 taaactgctgaaataaagtcttttcaaaaatcgtccttgtaatactcttcttcttcttcttcttcttct 3537
Db 3478 TAAACTTGCTGAAATTAAGTTTTTCAAAATCTGCTCTGTAATTAATTTCTTTCTTACA 3537
QY 3538 gtcttgccatactatacaactctgattcttcttcttcttcttcttcttcttcttcttcttcttcttct 3582
Db 3538 GTGCTTGCCATACTATATCAACTTTGATTCTTTGTACAACTTT 3582

RESULT 2
AF103908 5435 bp DNA PRI 14-AUG-2000
LOCUS Homo sapiens non-coding RNA DD3 gene, exons 2, 3, and 4.
DEFINITION AF103908
ACCESSION AF103908.1 GI:6165974
VERSION AF103908.1
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5435)
AUTHORS Bussemakers,M.J., van Bokhoven,A., Verhaegh,G.W., Smit,F.P.,
Karthaas,H.F., Schalken,J.A., Debruyne,F.M., Ru,N. and Isaacs,W.B.
DD3: a new prostate-specific gene, highly overexpressed in prostate
cancer
JOURNAL Cancer Res. 59 (23), 5975-5979 (1999)
MEDLINE 20072260
PUBMED 10606244
REFERENCE 2 (bases 1 to 5435)
AUTHORS Bussemakers,M.J.G., Van Bokhoven,A., Verhaegh,G.W., Smit,F.P.,
Karthaas,H.F.M., Schalken,J.A., Debruyne,F.M.J., Ru,N. and
Isaacs,W.B.
Direct Submission
JOURNAL Submitted (28-OCT-1998) Urology Research Laboratory, University
Hospital Nijmegen, P.O. Box 9101, Nijmegen 6500 HB, Netherlands
FEATURES
Source
1..5435
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/map="9q21-q22"
/repeat_region 1..78
/rpt_family="AluY"
/rpt_type=dispersed

```

```

exon 533..697
/note="alternative exon present in 5% of cDNA clones"
repeat_region 1035..1294
/rpt_family="Alu"
/rpt_type=dispersed
mRNA join(1571..1753,1981..5435)
/note="transcript III"
mRNA join(1571..1753,1981..2517)
/product="non-coding RNA DD3"
/note="transcript I"
mRNA join(1571..1753,1981..3579)
/product="non-coding RNA DD3"
/note="transcript (major) II"
exon 1571..1753
/number=3
exon 1981..5435
/number=4
polyA_signal 2495..2499
/note="transcript I"
polyA_site 2517
/note="transcript I"
polyA_signal 3553..3558
/note="transcript (major) II"
polyA_site 3579
/note="transcript (major) II"
repeat_region 5268..5423
/note="LINE"
polyA_site 5435
/rpt_family="L1"
/rpt_type=dispersed
BASE COUNT 1546 a 1188 c 1036 g 1654 t 11 others
ORIGIN
Query Match 84.7%; Score 3032.4; DB 88; Length 5435;
Best Local Similarity 99.5%; Pred No. 0;
Matches 3105; Conservative 0; Mismatches 11; Indels 6; Gaps 6;
QY 464 cagagatccctgggagaaatgccggccgccatcttgggtcatcgatgagctcgcctg 523
Db 1976 CACAGATCCTGGAGAAATGCCGGCCGCCATCTTGGGTCAATCGATGAGCTCGCCCTG 2035
QY 524 tccttggtccctgtgtgaggaagacattagaaaatgaattgattgttctctaaag 583
Db 2036 TGCTGTGCTCCCTGTGTGAGGAAGACATTAGAAATGAATGTGTCTCTTAAAGG 2095
QY 584 atgggcagaaacagatcctgtgtggatatttgaacgggattacagatttgaa 643
Db 2096 ATGGGCAGAAACAGATCCTGTGTGGATATTTATTGAACGGGATTACAGATTTGAAA 2155
QY 644 tgaagtcaaaagttagcattaccatgagagaaacagacagaaaatcttgatgct 703
Db 2156 TGAAGTCAAAAGTGAGCATTTACCAATGAGAGAAACAGACAGAGAAATCTTGATGGCT 2215
QY 704 tcacaagacatgcaacaaataatggaatactgtatgacatgagcagcagcaagctgg 763
Db 2216 TCACAAGACATGCACAAACAAATGGAATGATGTGATGATGACATGAGCAGCCAGCTGGG 2275
QY 764 gaggagataaccacggggcagaggggtcaggattcttgccctgctgcctaaactgtgcgtt 823
Db 2276 GAGGAGATAACACAGGGGAGAGGGTTCAGGATTTCTGCCCTGCTGCTAACTGTGCGTT 2335
QY 824 cataaccataatcatttcatttcttaacctcaaaacaaagctgttgaatatcgtatct 883
Db 2336 CATAACCAATCATTTTCTAATTTCTAACCTCAAAACAAAGCTGTGTGTAATATCTGATCT 2395
QY 884 ctacggttctctggggcccaacattctccatatatccagccacacactcatttttaatt 943
Db 2396 CTACGGTTCCTTCTGGGGCCCAACATTTCTCCATATATATCCAGCCACACACTCATTTTAAATTT 2455

```


Db 4614 ACATTAGATATTTAAAGCTCAAGAGGTTCAAAATCCAACTCATTAATCTCTCTTTCTTC 4673

Qy 3164 acctccctctctctccctatactactgattg-actgaacaggatgggtcccaaa-gat 3221

Db 4674 ACCT-CQCTGCTCTCTCCCTATATTAATGATGCACTGAACAGCATGGTCCCAATGTA 4732

Qy 3222 gccagtcacaaatgagaaacccagtggtccctctgttgatcatgcatgcaagactgctgaagc 3281

Db 4733 GCCATGCAAAATGAGAAACCCAGTGGCTCTTGTGGTACATGTCATGCAAGACTGCTGAAGC 4792

Qy 3282 cag-aggatgactgattacacccctcctggtgagggggaccactcctgggccttcggtatt 3340

Db 4793 CAGAAGATGACATGATTAACCCCTCATGTTGGTGGAGGGACACATCTCTGGCGTATT 4852

Qy 3341 gtcaggagcaagcctgagatgctccctgctcctcagtgctcctctgcatctccctcttcta 3400

Db 4853 GTCAGGAGCAAGACCTGAGATGCTCCCTGCTTTCAGTGTCTCTGTCATCTCCCTTTCTA 4912

Qy 3401 atgaagatccatagaatttctcatatttggaattccaaataggaactcacatgttttat 3460

Db 4913 ATGAAGATCCATAGAATTTGCTACATTTGAGAAATTCCAATTAGGAACCTCACATGTTTAT 4972

Qy 3461 ctccctcatcaatttttaaaactgctgaaataaagtgttttttcaaaaactgctccttgta 3520

Db 4973 CTGCCCTATCAATTTTAAACTTGCTGAAATTAAGTTTTCAAAATCTGCTCTGTGA 5032

Qy 3521 aattacttttctacagtgcttggcactactatatacaacttgatcttctgttacaact 3580

Db 5033 AATTACTTTTCTTACAGTGTCTTGCCATACTATATCAACTTTGATCTTTGTTACAAC 5092

Qy 3581 tt 3582

Db 5093 TT 5094

RESULT 3

AL390239

LOCUS Homo sapiens chromosome 9 clone RP11-58J3, *** SEQUENCING IN 09-MAR-2001

DEFINITION PROGRESS ***, 4 unordered pieces.

ACCESSION AL390239

VERSION AL390239.11 GI:13274794

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 164371)

AUTHORS Burton,J.

TITLE Direct Submission

JOURNAL Submitted (05-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk requests: clonerequest@sanger.ac.uk

COMMENT On Mar 12, 2001 this sequence version replaced gi:13273805.

----- Genome Center

Center: Sanger Centre

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: bA58J3

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 16337 bases at least Q40

Consensus quality: 163481 bases at least Q30

Consensus quality: 163577 bases at least Q20

Insert size: 164071; sum-of-ctg

Insert size: 166918; 4.2% error; agarose-fp

Quality coverage: 8.78x in Q20 bases; sum-of-ctgs Quality coverage: 9.07x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence recorded is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 20240: contig of 20240 bp in length

* 20241 20340: gap of 100 bp

* 20341 116100: contig of 95760 bp in length

* 116101 116200: gap of 100 bp

* 116201 118243: contig of 2043 bp in length

* 118244 118343: gap of 100 bp

* 118344 164371: contig of 46028 bp in length.

FEATURES

Location/Qualifiers

1..164371

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="9"

/clone="RP11-58J3"

/clone_lib="RPC1-11.1"

1..20240

/note="assembly_fragment:00838"

fragment_chain:1

clone_end:SP6

vector_side:left

20341..116100

/note="assembly_fragment:01998"

fragment_chain:1

116201..118243

/note="assembly_fragment:00916"

fragment_chain:1

118344..164371

/note="assembly_fragment:00334"

fragment_chain:1

clone_end:T7

vector_side:right

BASE COUNT 50254 a 34905 c 32663 g 46249 t 300 others

ORIGIN

Query Match 82.6%; Score 2959.8; DB 80; Length 164371;

Best Local Similarity 99.2%; Pred. No. 0;

Matches 3100; Conservative 0; Mismatches 12; Indels 13; Gaps 12;

Qy 464 cagagatccctgggagaaaatgccggcccccattctgggtcatcgatgagctgcgcctg 523

Db 33316 CACAGATCCCTGGGAGAAATGCCGGCCGCCATCTTGGGTGTCATGCGATGAGCCTCGCCCTG 33375

Qy 524 tgctggtccctgtgaggaagacattagaaaatgaattgattgttctctaaagg 583

Db 33376 TGCCTGGTCCCGCTTGTGAGGAAGACATTAAGAAATGAATGATGTTCTTCTTAAGG 33435

Qy 584 atgggcaggaaacagatcctgttggtgatatatttgaacgggattacagatttgaa 643

Db 33436 ATGGGCAGGAAACAGATCCTGTGTGGATATTTATTTCACGGGATTACAGATTGAAA 33495

Qy 644 tgaatcacaagtgaagcattaccatgagaggaagaaacagagagaaaattcttgatgct 703

Db 33496 TGAAGTCACAAAGTGAGCATTTACCAATGAGAGGAAACAGACGAGAAAATCTTGATGGCT 33555

Qy 704 tcacaagacatcaacaaacaaatggaatactgtgatgacatgagcgagcgaagctggg 763

Db 33556 TCACAAGACATGCACAAACAAATGGAATAGTGTGATGACATGAGCGACGCAAGCTGGG 33615

Qy 764 gaggagataaccacggggcagagggtcaggattcttggccctgctgctctaaactgtgcgt 823

Db 33616 GAGGAGATAACACACGGGGCAGAGGGTTCAGGATTCCTGGCCCTCTGCTCTAACTGTCGTT 33675

Qy 824 cataaccaaatcatattcatatttctaaccctcaaaacaaagctgtgttaatactgatct 883

Db 33676 CATAACCAAAATCATTTCAATTTTCTAAACCTCAAAACAAAGCTGTTGTAATATCTGATCT 33735

Qy	884	ctacggttcctctcgggcccacacattctccatatataccagccacactcatttttbaattt	943
Db	33736	ctacgggttcctctcgggcccacacattctccatatataccagccacactcatttttbaattt	33795
Qy	944	tagttccagatctgtagcttgacacttctacactgtagaataacatttactcatttggtt	1003
Db	33796	tagttccagatctgtagcttgacacttctacactgtagaataacatttactcatttggtt	33855
Qy	1004	caaaagacccttcgttgcctcctaataatgtagctgactgtttttcctaaagagtgctct	1063
Db	33856	caaaagacccttcgttgcctcctaataatgtagctgactgtttttcctaaagagtgctct	33915
Qy	1064	ggcccaagggaatctgtgaacaaggctggaagcactctcaagatctcttcaggggttatactt	1123
Db	33916	ggcccaagggaatctgtgaacaaggctggaagcactctcaagatctcttcaggggttatactt	33975
Qy	1124	actagcacacagcatgatctacgagtgaaattactaaaccaacatcatcctcagtgctc	1183
Db	33976	actagcacacagcatgatctacgagtgaaattactaaaccaacatcatcctcagtgctc	34035
Qy	1184	tttgccatactgaaattcatttcccacttttgcgccattctcaagacctcaaaaatgctc	1243
Db	34036	tttgccatactgaaattcatttcccacttttgcgccattctcaagacctcaaaaatgctc	34095
Qy	1244	attccattaatcacaggattaaacttttttttaacctgggaagaattcaattgttacct	1303
Db	34096	attccattaatcacaggattaaacttttttttaacctgggaagaattcaattgttacct	34155
Qy	1304	gcagctatgggaatttaattacattttgtttccagtgcaaaagatgaactaaagtccttt	1363
Db	34156	gcagctatgggaatttaattacattttgtttccagtgcaaaagatgaactaaagtccttt	34215
Qy	1364	atccctccctttgttgatttttttccagtaataaagttaaaatgcttccttgtaact	1423
Db	34216	atccctccctttgttgatttttttccagtaataaagttaaaatgcttccttgtaact	34275
Qy	1424	gagctgtatacag-cacagcctctcccccctccctcagccttatctgtcataccatca	1482
Db	34276	gagctgtatpacagccacacagcctctcccccctccctcagccttatctgtcataccatca	34335
Qy	1483	accctcccataccacacctaaacaaaacttaacttgtaattccttggaacatgtaggacat	1542
Db	34336	accctcccataccacacctaaacaaaacttaacttgtaattccttggaacatgtaggacat	34393
Qy	1543	acattattcctctgcctgagaagcttccctgtctcttaaatctagaatgtagtaaaag	1602
Db	34394	acattattcctctgcctgagaagcttccctgtctcttaaatctagaatgtagtaaaag	34453
Qy	1603	tttggaaatgaattgactatcttactctcatgcaaaagaggacacatatgagatcatcat	1662
Db	34454	tttggaaatgaattgactatcttactctcatgcaaaagaggacacatatgagatcatcat	34513
Qy	1663	cacatgacagcgaataactaaaagtgtaatttgattataagagtttagataaaatatag	1722
Db	34514	cacatgacagcgaataactaaaagtgtaatttgattataagagtttagataaaatatag	34573
Qy	1723	aaatgcaagaccacagaggaattgtttatgggcacgctttgtaagcctggagtgtaag	1782
Db	34574	aaatgcaagaccacagaggaattgtttatgggcacgctttgtaagcctggagtgtaag	34633
Qy	1783	caagaggcagggaacctcatgattcttataaaatacttacttctctactctatacacc	1842
Db	34634	caagaggcagggaacctcatgattcttataaaatacttacttctctactctatacacc	34693
Qy	1843	aataaccaagaagcttttcacagaattcagtcagtcaggaatcccccaaggtaaaccttat	1902
Db	34694	aataaccaagaagcttttttcacagaattcagtcagtcaggaatcccccaaggtaaaccttat	34753
Qy	1903	ccatttcaggtgtagcgccttttagaattttggcaaatcatacttggtcaccttatctcaac	1962
Db	34754	ccatttcaggtgtagcgccttttagaattttggcaaatcatacttggtcaccttatctcaac	34813

Qy	1963	t t g a g a t g t g t t g t c c t t g t a g t t a a t t g a a a g a a t a g g c a c t c t t t g t a g c c a c t	2020
Db	34814	T T T G A G A T G T G T T G C C T T G T A G T T A A T T G A A A A A A T A G G C A C T C T T G T G A G C C A C T	34873
Qy	2023	t t a g g t t c a c t c c t g g a a t a a a a t t t a c a a g a g c t a c t c a g g a c c a g t t a a g	2082
Db	34874	T T A G G G T T C A C T C C T G G C A A T A A A G A A T T A C A A G A G C T A C T C A G G A C C A G T T G T T A A G	34933
Qy	2083	a g c t c t g t g t g t g t g t g t g t g a g t g t a c a t g c c a a a a g t g t g c c t c t c t c -	2141
Db	34934	A G C T C - - T G T G T G T G T G T G T G T G A G T G A C A T G C C A A A G T G C C T C T C T C T C T	34991
Qy	2142	t b g a c c a t t a t t c a g a c t t - a a a c a a g c a t g t t t t c a a t g g c a c t a g a g c t g c c a	2200
Db	34992	T T G A C C A T T A T T T C A G A C T T A A A A C A A G C A T G T T T C A A A T G C A C T A T A G A G T G C C A	35051
Qy	2201	a t g a t g t a t c a c c a c a t a t c a t c a t t a t t c c a g t a a a t g t g a t a a a t g t c a t e t g t	2260
Db	35052	A T G A T G T A T C A C C A C A T A T C A T A T T T C C A G T A A A T G A A T A A T G T G C A T C T G T	35111
Qy	2261	t a a c a t a a a a a a g t t g a c t t c a a a a g c a g t g g a a t g g a a t g g a c c a c c a a t a t g a t	2320
Db	35112	T A A C A T A A A A A A G T T G A C T T C A C A A A G C A G C T G G A A A T G S A C A C C A C A A T A T G C A T	35171
Qy	2321	a a a t c a a c t c t a c a c a t a g a c a c a c t g t c t g a c a t a t a t t g t a g a a g c a c c t g c	2380
Db	35172	A A A T C T A A C T C T A C A T C A G C T A C A C A C T G C T T G A C A T A T A T T G T A G A A G C A C T G C	35231
Qy	2381	a t t t g g g t c t c t t a a g c a a a a c t i g c a t t a g g t c a g t c a g t g g g c t g t g a t c a g	2440
Db	35232	A T T T G T G G G T C T T T A G C A A A A T A C T T G C A A T A G G T C A G C T G G G C T G T G C A T C A G	35291
Qy	2441	g c g t t t g g a a a t t c a a t c t c t c a g a a g c c a g a a t t t g a a t c c c t c a t c t t t a	2500
Db	35292	G C G T T T G A G A A A T A T C A A T T C T C A G C A A G C C A G A A T T G A A T C C C T C A T C T T T A	35351
Qy	2501	g g a a o a t t a c c a g g t t g g a g a g a t t c a g a c a g c t o a g t g t t t o a c t a a t g t c t c	2560
Db	35352	G G A A T C A T T A C A C A G T T T G G A G A G A T T C A G A C A G C T C A G C T G C T T C A C T A A A T G T C T C	35411
Qy	2561	t g a a c t c t g t c c c t c t t g t g t c a t g a t a g t c a a t a a a t a a t g t a t c t t t g a a c t	2620
Db	35412	T G A A C T C T G T C C C T T T G T G T T C A T G G A T A G T C C A A T A A T A A T G T T A T C T T T G A C T	35471
Qy	2621	g a t g t c a t a g g a g a a t a a g a a c t c t g a t a t c a a c a t a g g a t t c a a a g a a	2680
Db	35472	G A T G C T A T A G G A G A A T A A A G A A C T C T G A G T G A T A T C A C A T T A G G G A T T C A A A G A	35531
Qy	2681	a t a t a g a t t a a g t c a c a c t g t t c a a a g g a c c a a g a t a c a a g a a c t c t g a c t g t	2740
Db	35532	A A T A T A G A T T A A A G T C A C A C T G G T C A A A G G A A C A A G A T A C A A A A A C T C T G A G C T G T	35591
Qy	2741	c a t c t c c c a t c t g t g a g c c a c a c c a a c a g a c c a a c a g c a t g t c t g a g a t c c	2800
Db	35592	C A T C G T C C C A T C T C T G T A G C C A C A C C A C A G A G G A C C C A A C G C A T G T C T G A T A T C C	35651
Qy	2801	t t a a a t c a a g a a a c c a g t g t c a t g a t t g a a t t c c t a t t a t g a t g t a g c t c t g g	2860
Db	35652	T T A A A T C A A G A A A C A C A G T C A T G A G T G A A T T C C T A T T A T G A T G C T A G C T T C T G G	35711
Qy	2861	c a a t c t c t g g c t c c t c t t g a c a c a t a t t a g c t c t a g c c t t g c t c c a c g a c t t t a	2920
Db	35712	C C A T C T C T G C C T C C T C T T G A C A C A T A T A G C T T C T A G C T T T G C T C C A G A C T T T T A	35771
Qy	2921	t c t t t t c c c a c a c a t c g c t t a c c a a c t c t c t c t c t c t g t t g c t t t g g a c t t c c c c	2980
Db	35772	T C T T T T T C C A A C A C A T C G C T T A C C A A T C C T C T C T G C T T G T T G C T T T G A C T T C C C C	35831
Qy	2981	a c a a g a a t t c a a c g a c t c a a g t c t t t t c t c a t c c c a c c a c t a a c c t g a a t g c c	3040
Db	35832	A C A A G A A T T C A A C A G A C T C T A A G T H C T T T T C T T C C A T C C C C A C C A C T A A C C T G A A - T G C C	35890
Qy	3041	t a g a c c c t a t t t t a t a a t t c c a a t a g a t g t g c g c t a t g g g c a a t a t g t c t t a g a	3100

```
Db 35891 TAGACCTTATTTTATTAATTTCCCAATAGATGCTGCTATGGCT-ATATGCTTTAGA 35949
QY 3101 tgaacattagatattaaagtcaagaggttcaaaatcccaactcattatctctcttct 3160
Db 35950 TGAACATTAGATATTTAAAGCTCAAGAGGTTCAAAATCCAACTCATTTCTCTTCT 36009
QY 3161 ttcaactccctgctctctccctatatattactgattg-actgaacagatggtcccaaa- 3218
Db 36010 TTCACT-CCCTGCTCTCTCCCTATATTTACTGATTGCTCACTGAACAGCATGTTCCCAAT 36068
QY 3219 gatccagtcacaaatgagaaacccagtggtcctctgtggtatcgtcgtcgaagactgctga 3278
Db 36069 GTAGCCATGCAAAATGAGAAACCCAGTGGCTCTGTGTGTATGCTATGCAAGACTGCTGA 36128
QY 3279 agccag-agcatgactgattacccctcatggtggaaggaacacactcctggccttcgtg 3337
Db 36129 AGCAGAAGGATGACTGATTACGCCTCATGGGTGAGGGGACCACCTCCTGGGCTTCGTG 36188
QY 3338 attgtcaggagcaagactgagatgctccctgctcctcagtgctcctgcatctcccttt 3397
Db 36189 ATTGTGAGGAGCAAGACTGAGATGCTCCCTGCTTCCAGTGTCTCTGCTGCTCTCCCTTT 36248
QY 3398 ctaatgaagatcatagaaattgctacatttggaattccaattaggaactcacaatgttt 3457
Db 36249 CTAATGAAGATCCATAGAATTTGCTACATTTGAGAAFTCCAATTAGGAACACATGTTT 36308
QY 3458 tatctgctctatacaatttttaactgctgaaattaaagtatttttcaaaatcgtcctt 3517
Db 36309 TAATGCTCCCTATCAATTTTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 36368
QY 3518 gtaaatctctttcttacagtgctcgtggtacatactatcaaatctgattcttctgttaca 3577
Db 36369 GTAAATTAATTTTCTTACAGTGTCTTGGCATACTATATCAACTTTGATTTGTTTACA 36428
QY 3578 acttt 3582
Db 36429 ACTTT 36433

RESULT 4
AL359314
LOCUS
DEFINITION Homo sapiens chromosome 9 clone RP11-108L4, *** SEQUENCING IN
ACCESSION AL359314
VERSION AL359314.12 GI:13396560
KEYWORDS HTG: HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 173831)
Leongamornlert.D.
Direct Submission
Submitted (07-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Mar 20, 2001 this sequence version replaced gi:13277120.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA108L4
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 173545 bases at least Q40
Consensus quality: 173599 bases at least Q30
Consensus quality: 173621 bases at least Q20
```

```
Insert size: 173631; sum-of-contigs
Insert size: 172123; 10.0% error; agarose-fp
Quality coverage: 10.35x in Q20 bases; sum-of-contigs Quality
coverage: 10.50x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```

```
* 1 50595: contig of 50595 bp in length
* 50596 50695: gap of 100 bp
* 50696 157636: contig of 106941 bp in length
* 157637 157736: gap of 100 bp
* 157737 173831: contig of 16095 bp in length.
```

FEATURES

source

```
1..173831
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone_lib="RPC1-11.1"
1..50595
```

misc_feature

```
1..50595
/fragment="assembly fragment:04137"
fragment_chain:1
```

misc_feature

```
50696..157636
/fragment="assembly fragment:04236"
fragment_chain:1
```

misc_feature

```
157737..173831
/fragment="assembly fragment:01656"
```

BASE COUNT 52107 a 36469 c 35018 g 50037 t 200 others

ORIGIN

```
Query Match 38.1%; Score 1363.4; DB 80; Length 173831;
Best Local Similarity 92.8%; Pred. No. 0;
Matches 1400; Conservative 0; Mismatches 106; Indels 3; Gaps 3;

QY 464 cagagatccctgggagaaatgccggccgccatcttgggtcatcgatgcctgcctcg 523
Db 156237 CACAGATCCTGGGAGAAATGCCGCCGCCATCTTGGGTCTCGATGAGCTCGCCCTG 156296
QY 524 tccttggtccctctgtgaggaagacattagaaaataaattgattgttcttaaaag 583
Db 156297 TGCTGTGCTCCCTTGTGAGGAAGGACATTAGAAATGAATGTGTCTCTTAAAGG 156356
QY 584 atgggcaggaacacagatcctgtgtggatattttgaacgggattacacatttgaaa 643
Db 156357 ATGGGCAGGAAACACAGATCCTGTGTGGGATATTTATTTCACGGGATTACAGATTTGAAA 156416
QY 644 tgaagtcaaaagtgaagcattaccatgagaggaacacagagaaaaactcttgatg 703
Db 156417 TGAAGTCACAAAGTGAAGCATTTACCAATGAGAGGAAACACAGAGAAAAATCTTGATGGCT 156476
QY 704 tcacaagacatgcaacaaacaaatggaatactgtgatacagatgagcgagcaagctgg 763
Db 156477 TCACAGACATGCACAAACAAATGGAATACTGTGATGATGAGGACGAGCAAGCTGGG 156536
QY 764 gaggagataaacacggggcagaggggtcaggattcttggcctgctgcctaaactgtgcgtt 823
Db 156537 GAGGAGATAAACACGGGCGAGAGGGTTCAGGATTCCTGGCCCTCTGCTCTAACTGTGCGTT 156596
QY 824 cataaccataatcatttcatttcttaacccctcaaaacaaagctgtgtgtaatatctgatct 883
Db 156597 CATAACCAAAATCATTTCAATATTTTCTTAACCCCTCAAAACAAAGCTGTGTGTAATATCTGATCT 156656
QY 884 ctacagttcctctggggcccaacattctccatatatccagccacacactcatttttaatt 943
Db 156657 CTACGGGTTCTTCTGGGGCCCAACATTTCTCCATATATATCCAGCCACACTCATTTTAAATTT 156716
```

```
QY 944 tagttcccagatctgtactgtgacacctttctacactgtagaataaacattactcttggtt 1003
|||||
Db 156717 TAGTTCCTCCAGATCTGTACTGTGACCTTTCTACACTGTAGAAATAACATTACTATTGTTG 156776
QY 1004 caaagacccttggttgctgctcctaataatgttagctgactgtttttctcctaagagagtgctt 1063
|||||
Db 156777 CAAAGACCCTTCGTGTGCTGCTTAATATGTAGCTGACTGTTTTCCTTAAGGAGTGTTCT 156836
QY 1064 gcccaggggactctgtaacagctgggaagcatctcaagatcttccagaggttatactt 1123
|||||
Db 156837 GCCCAGGGGATCTGTGACAGCGCTGGGAAGCATCTCAAGATCTTTCAGAGGTTATAC 156896
QY 1124 actagcacacagcatgatcattacagggagtggaattatctaatcaaacatcatctcagtgctc 1183
|||||
Db 156897 ACTAGCACACAGCATGATCATTTACGGAGTGAATTATCTAATCAACATCATCTCTCAGTGTC 156956
QY 1184 ttggccataactgaattcatttcccacttttggccatttctcgaagcctcaaaatgctc 1243
|||||
Db 156957 TTGGCCCATACTGAAATTCATTTCCCACTTTTGTGCCCAATCTCAAGACCTCAAAATGTC 157016
QY 1244 attccattaataatcacagattaaacttttttttaacctggaagaattcaatgtttacat 1303
|||||
Db 157017 ATTCCATTAATATCAGAGATTAACCTTTTTTTTAACTGGAAGATTCATGTTTACAT 157076
QY 1304 gcagctatgggaatttaattacatatatttggttttccagtgcaaatgactaagtccttt 1363
|||||
Db 157077 GCAGCTATGGGAATTTAATTACATATTTGTTTTCAGTGCAGAGATGACTAAGTCCTTT 157136
QY 1364 atccctccccttggttgatttttttccagttataaagttaaaatgcttagcctgtact 1423
|||||
Db 157137 ATTCCTCCCTTTGTTGATTTTTTTTCCAGTATAAAGTTAAATGCTTAGCTTTGCTACT 157196
QY 1424 gaggctgtatacag-cacagcctctcccatccctccagccttactctgtcatcacatca 1482
|||||
Db 157197 GAGGCTGTATACAGCCAGAGCTCTCCCACTCCCTCCAGCCCTTATCTGTCTACACATCA 157256
QY 1483 acccttcccataccactcaaaactaaacttgtaattccttgaacatgtcaggacat 1542
|||||
Db 157257 ACCCTCCCAT-GCACCTAAACAAATCTAAGTTGTAATTCCTTGAACATGTCAGG-CAT 157314
QY 1543 acattattccttctgctgagagctctccttctgtctcttcttaactagatgataag 1602
|||||
Db 157315 ACATTATTCTCTTGCTGCTGAGAGCTCTTCTCTGTCTCTTAAATCTAGAAATGATGTAAG 157374
QY 1603 ttttgaataagtgtactatcttactctgcaagaaggacacatatagattcatcat 1662
|||||
Db 157375 TTTTGAATPAAGTTGACTATCTTACTTCTATGCAAGAGGGACACATATGAGATTCATCAT 157434
QY 1663 cacatgagacagcaataactaaaagtgaattttgattataaagagtttagataaaatatg 1722
|||||
Db 157435 CACATGAGACAGCAAAATACTAAAAGTGTAATTTGATTATAAGAGTTTACATAAATATATG 157494
QY 1723 aaatgcaagagccacagaggaatgtttatggggcaactgtttagcctggagtggaag 1782
|||||
Db 157495 AAATGCCAAGAGCCACAGAGGGAAATGTTTATGGGCGACGTTTGTAAAGCCCTGGATGTAAG 157554
QY 1783 caaagcagggaaacctcatgtatcttataataataacttcttctctatctctatcac 1842
|||||
Db 157555 CAAAGCAGGGAAACCTCATCATATCTATATATATATATCTTCTATCTCTATCTATCAC 157614
QY 1843 aatatcaacaagctttttcacagaaattcatgcatgcaaatcccccaaggtaacctttat 1902
|||||
Db 157615 AATATCCAACAAGCTTTTTCACANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 157674
QY 1903 ccatttcatgggtgagtcgcttagaattttggcaaatcatactggtcattctatcacaac 1962
|||||
Db 157675 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 157734
QY 1963 tttagatg 1971
|||||
Db 157735 NNNNACAAG 157743
```

RESULT 5
AL358573/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

AL358573 267581 bp DNA HTG 15-APR-2001
Homo sapiens chromosome 9 clone RP11-133022, *** SEQUENCING IN
PROGRESS ***, 37 unordered pieces.
AL358573
AL358573.17 GI:13660951
HTG: HTGS_PHASE1; HTGS_DRAFT.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
McLay, K.
Direct Submission
Submitted (14-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Apr 17, 2001 this sequence version replaced gi:13398774.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bal33022
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 256699 bases at least Q40
Consensus quality: 259744 bases at least Q30
Consensus quality: 261407 bases at least Q20
Insert size: 263881; sum-of-contigs
Insert size: 135491; 19.3% error; agarose-fp
Quality coverage: 5.51x in Q20 bases; sum-of-contigs Quality
coverage: 12.23x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 25718: contig of 25718 bp in length
* 25719 25818: gap of 100 bp
* 25819 29567: contig of 3749 bp in length
* 29568 29667: gap of 100 bp
* 29668 40520: contig of 10853 bp in length
* 40521 40620: gap of 100 bp
* 40621 50002: contig of 9382 bp in length
* 50003 50102: gap of 100 bp
* 50103 54942: contig of 4840 bp in length
* 54943 55042: gap of 100 bp
* 55043 59067: contig of 4025 bp in length
* 59068 59167: gap of 100 bp
* 59168 72327: contig of 13160 bp in length
* 72328 72427: gap of 100 bp
* 72428 79396: contig of 6969 bp in length
* 79397 79496: gap of 100 bp
* 79497 82738: contig of 3242 bp in length
* 82739 82838: gap of 100 bp
* 82839 87543: contig of 4705 bp in length
* 87544 87643: gap of 100 bp
* 87644 89847: contig of 2204 bp in length
* 89848 89947: gap of 100 bp
* 89948 92376: contig of 2429 bp in length
* 92377 92476: gap of 100 bp
* 92477 99799: contig of 7323 bp in length
* 99800 99899: gap of 100 bp
* 99900 102117: contig of 2218 bp in length

```
* 102118 102217: gap of 100 bp
* 102218 109330: contig of 7113 bp in length
* 109331 109430: gap of 100 bp
* 109431 112187: contig of 2757 bp in length
* 112188 112287: gap of 100 bp
* 112288 116407: contig of 4120 bp in length
* 116408 116507: gap of 100 bp
* 116508 119386: contig of 2879 bp in length
* 119387 119486: gap of 100 bp
* 119487 121869: contig of 2383 bp in length
* 121870 121969: gap of 100 bp
* 121970 125546: contig of 3577 bp in length
* 125547 125646: gap of 100 bp
* 125647 128990: contig of 3344 bp in length
* 128991 129090: gap of 100 bp
* 129091 134055: contig of 4965 bp in length
* 134056 134155: gap of 100 bp
* 134156 138314: contig of 4159 bp in length
* 138315 138414: gap of 100 bp
* 138415 140612: contig of 2198 bp in length
* 140613 140712: gap of 100 bp
* 140713 143296: contig of 2584 bp in length
* 143297 143396: gap of 100 bp
* 143397 148113: contig of 4717 bp in length
* 148114 148213: gap of 100 bp
* 148214 151009: contig of 2796 bp in length
* 151010 151109: gap of 100 bp
* 151110 156616: contig of 5507 bp in length
* 156617 156716: gap of 100 bp
* 156717 158763: contig of 2047 bp in length
* 158764 158863: gap of 100 bp
* 158864 162129: contig of 3266 bp in length
* 162130 162229: gap of 100 bp
* 162230 164443: contig of 2214 bp in length
* 164444 164543: gap of 100 bp
* 164544 169800: contig of 5257 bp in length
* 169801 169900: gap of 100 bp
* 169901 251545: contig of 81645 bp in length
* 251546 251645: gap of 100 bp
* 251646 259029: contig of 7384 bp in length
* 259030 259129: gap of 100 bp
* 259130 262871: contig of 3742 bp in length
* 262872 262971: gap of 100 bp
* 262972 265209: contig of 2238 bp in length
* 265210 265309: gap of 100 bp
* 265310 267581: contig of 2272 bp in length.
```

FEATURES

source

```
1..267581
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-133022"
/clone_lib="RPC1-11.1"

misc_feature 1..25718
/note="assembly_fragment:00645"
fragment_chain:1
clone_end:r7
vector_side:left"

misc_feature 25819..29567
/note="assembly_fragment:01660"
fragment_chain:1"

misc_feature 29668..40520
/note="assembly_fragment:03080"
fragment_chain:1"

misc_feature 40621..50002
/note="assembly_fragment:00771"
fragment_chain:1"

misc_feature 50103..54942
/note="assembly_fragment:02629"
fragment_chain:1"

misc_feature 55043..59067
/note="assembly_fragment:04565"
fragment_chain:1"
```

```
misc_feature 59168..72327
/note="assembly_fragment:00223"
fragment_chain:1"

misc_feature 72428..79396
/note="assembly_fragment:03318"
fragment_chain:1"

misc_feature 79497..82738
/note="assembly_fragment:00224"
fragment_chain:2"

misc_feature 82839..87543
/note="assembly_fragment:03050"
fragment_chain:2"

misc_feature 87644..89847
/note="assembly_fragment:01948"
fragment_chain:3"

misc_feature 89948..92376
/note="assembly_fragment:02191"
fragment_chain:3"

misc_feature 92477..99799
/note="assembly_fragment:00010"

misc_feature 99900..102117
/note="assembly_fragment:00672"

misc_feature 102218..109330
/note="assembly_fragment:01326"

misc_feature 109431..112187
/note="assembly_fragment:01498"

misc_feature 112288..116407
/note="assembly_fragment:01607"

misc_feature 116508..119386
/note="assembly_fragment:01899"

misc_feature 119487..121869
/note="assembly_fragment:02068"

misc_feature 121970..125546
/note="assembly_fragment:02469"

misc_feature 125647..128990
/note="assembly_fragment:02502"

misc_feature 129091..134055
/note="assembly_fragment:02669"

misc_feature 134156..138314
/note="assembly_fragment:02757"

misc_feature 138415..140612
/note="assembly_fragment:02776"

misc_feature 140713..143296
/note="assembly_fragment:03127"

misc_feature 143397..148113
/note="assembly_fragment:03634"

misc_feature 148214..151009
/note="assembly_fragment:03898"

misc_feature 151110..156616
/note="assembly_fragment:04220"

misc_feature 156717..158763
/note="assembly_fragment:04276"

misc_feature 158864..162129
/note="assembly_fragment:04418"

misc_feature 162230..164443
/note="assembly_fragment:04428"
```

```
Query Match 36.8%; Score 1319.8; DB 80; Length 267581;
Best Local Similarity 92.2%; Pred.No. 0;
Matches 1360; Conservative 0; Mismatches 112; Indels 3; Gaps 3;
```

```
QY 464 cagagatccctgggagaaatgccggccgcacattctgggtcatcgatgagcttcgccclg 523
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165900 CACAGATCCCTGGGAGAAATGCCCGCCGCCATCTTGGGTCAATCGATGAGCTTCGCCCTG 165841

QY 524 tgctgtgctccctgtgaggggaagacattagaaaaatgaattgattgttctcctaaag 583
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165840 TGCCTGTGCTCCGCTTGTGAGGGAAGGACATAGAAAATGAATTGATGTGTTCTTTAAAG 165781

QY 584 atgggcagaaacacatcctgtgtggtattattttaaaggattacagatttgaaa 643
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165780 ATGGCGAGGAAACAGATCCTGTGTGGATATTTATTTGAACGGGATTACAGATTTGAAA 165721
```

QY 644 tgaagtcaaaagttagcattaccattgagaggaagaaacagagcaagaaatcttgatgct 703
|||||
Db 165720 TGAAGTCACAAAGTCAGCATTTACCAATGAGAGGAAACAGAGCAAAATCTTGATGCT 165661
|||||
QY 704 tcaagacatgcaacaaacaaaaatggaactgtgtagacatgaggcagcaagctggg 763
|||||
Db 165660 TCACAAGACATGCAACAACAAAATGGAATACATGTGATGACATGAGGAGCAAGCTGG 165601
|||||
QY 764 gaggagataaacacaggggcaagaggtcaggattctggccctgcctcaaacctgagct 823
|||||
Db 165600 GAGGAGATAAACCCGGGAGAGGCTCAGGATCTGGCCCTGCTGCCATAAATGTGCGGT 165541
|||||
QY 824 cataaaccattctcatattcttaacccctcaaaacaaagctgtgtaatatctgatct 883
|||||
Db 165540 CATACCAAAATCATTTTCAATTTCTTAACCCTCAAAACAAAGCTGTGTAATATCTGATCT 165481
|||||
QY 884 ctacgggtctctctgggcccacatactctccatataatccagccacacactcatctttaaatt 943
|||||
Db 165480 CTACGGTTCCTTCTGGGCCCAACATTTCCATPATATCCAGCCACACTCATTTTAAATATT 165421
|||||
QY 944 tagttccagatctgactgtgacctttctacactgtagaataacattactatttctgt 1003
|||||
Db 165420 TAGTTCCAGATCTGTACTGTGACCTTTCTACACGTGAGATAACATTAATCTTTGTT 165361
|||||
QY 1004 caaagaccctctgtgtgctgctctaattgttagctgactgttttctcctaagagtgcttct 1063
|||||
Db 165360 CAAGACCTTCGCTGCTGCTTAATATGTAGCTGACTGTGTTTCTTAAGGAGGTCTT 165301
|||||
QY 1064 ggcacaggggatactgtgaacagggctgggaagcatctcaagatctctccaggggttaactt 1123
|||||
Db 165300 GGCCCAAGGGGATCTGTGAACAGGCTGGGAAGCATCTCAAGATCTTTCCAGGGTTATACTT 165241
|||||
QY 1124 actagcacacagcatgatcattacagagtggaattacttaatacaacatcatctcaagtgc 1183
|||||
Db 165240 ACTAGCACACAGCATGATCATTACGGAGTGAAATATCTAATCAACATCATCTCAGTGTC 165181
|||||
QY 1184 ttggccataactgaaatcattctccactttgtgccattctcaagacctcaaaatgctc 1243
|||||
Db 165180 TTTGCCATACAGAAATTCATTTCCTCCACTTTTGCCCATCTCAAGACCTCAAAATGTC 165121
|||||
QY 1244 attccatttaataacacagattaaacttttttttaacctggaagaatcaatgttaact 1303
|||||
Db 165120 ATTCCATTAAATATCAGAGATTAACTTTTTTTTAAACCTGGAAGAATTCATGTTTACAT 165061
|||||
QY 1304 cgactatgggaatttaattacattttgttttccagtgcaaaatgaactgaagctcttt 1363
|||||
Db 165060 GCAGCTATGGGAATTAATATACATTTTGTGTTCCAGTGCAAAAGATGACTAAGTCCCTT 165001
|||||
QY 1364 atccctccctctgtgtgatttttttccagataaaagttaaagtcttagccttgact 1423
|||||
Db 165000 ATCCCTCCCTTTGTTGATTTTTTTTCCAGTATAAGTTAAATGCTTAGCCTTGACT 164941
|||||
QY 1424 gaggctgtatacag-cacagcctctcccatccctccagccttatctgtcatcacata 1482
|||||
Db 164940 GAGGCTGTATAGAGCACACAGCCTCTCCCATCCCTCCAGCCTTATCTGTCTATCACCATCA 164881
|||||
QY 1483 accctcccatcacacacatacaaaatcaacttgaatttcccttaaaatctagaatgataag 1542
|||||
Db 164880 ACCCTCCCAT -GCACATAAACAAATCTPACTTGTGAATCTTGAACATGTCAGG-CAT 164823
|||||
QY 1543 acattattctctgctgagagctctctctgtctctttaaactagatgataag 1602
|||||
Db 164822 ACATPATCTCTGCTGAGAGCTCTTCCCTGTCTTAATCTAGAAATGATGTAAG 164763
|||||
QY 1603 ttttgaataagttgacttacttactcaatgcaaaagggagcacatatgagatcatcat 1662
|||||
Db 164762 TTTTGAATAAGTTGACTATCTTACTTTCATGCAAAAGGAGGACACATATGAGATTATCAT 164703
|||||
QY 1663 cacatgagacagcaataactaaaagtgttaatttgattataagatttagataaataatg 1722
|||||
Db 164702 CACATGAGACAGCAATACTAAAGTGAATTTGATTATTAAGAGTTTGTAGATAAATATAG 164643
|||||
QY 1723 aaatgcaagagccacagaggggaatgtttatg999gacgcttttgaagcctgggagtggaag 1782

|||||
Db 164642 AAATCAAGAGCCACAGAGGAATGTTTATGGGGCACCTTTGTAGCCTGGGATGTGAAG 164583
|||||
QY 1783 caaagcagggaacctatagatattatataataataacttctctctctctatctacac 1842
|||||
Db 164582 CAAAGCGAGGAACTCATAGTATCTTATATAATATATTTNNNNNNNNNNNNNNNN 164523
|||||
QY 1843 aatatccacaagcttttcacagaattcatgcagtgcaaatccccaaggtaaacctttat 1902
|||||
Db 164522 NNN 164463
|||||
QY 1903 ccattcatggtgagtcgctttagaatttgga 1937
|||||
Db 164462 NNN 164428
|||||
RESULT 6
G56925/c G56925 546 bp DNA STS 30-MAR-2000
LOCUS SHGC-102485 Human Homo sapiens STS genomic, sequence tagged site.
DEFINITION
ACCESSION G56925
VERSION G56925.1 GI:6122094
KEYWORDS STS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 546)
AUTHORS Olivier, M. and Cox, D.R.
TITLE Unpublished, Olivier, M., Cox, D.R. (2000)
JOURNAL Unpublished (2000)
COMMENT
Contact: Michael Olivier, David R. Cox
Stanford Human Genome Center
Stanford University School of Medicine
4005 Miranda Ave. 2nd Fl., Palo Alto, CA 94025, USA
Tel: (650) 320-5800
Fax: (650) 320-5801
Email: olivier@hgsc.stanford.edu
Primer A: TTAAGGATCTCAGACATGCGCTG
Primer B: TCATTACCAGGTTGGAGAGGA
STS size: 300
PCR Profile:
Initial incubation: 95 degrees C for 10 minutes
Denaturation: 94 degrees C for 30 seconds
Annealing: 60 degrees C for 30 seconds
Polymerization: 72 degrees C for 23 seconds
PCR Cycles: 30
Thermal Cycler: Perkin Elmer 9700
Protocol:
Template: 25 ng
each 1 uM
Primer: each 200 uM
dNTPs: Ampliqaq Gold Polymerase: 0.07 units/ul
Total Vol: 5 ul
Buffer:
MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 8.3
BAC ends sequenced at TIGR from the RPC111 BAC library. Designed and developed at the Stanford Human Genome Center.
FEATURES
source
Location/Qualifiers
1..546
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="9"
/clone_lib="Human"
29..328
primer_bind
29..51
primer_bind
complement(306..328)

QY 2783 aacgatgtctgagatccttaataaagaacaccagtgctcatgagtggaattc 2835
|||||
Db 62 AACGATGCTCAGATCCTTAATAAAGGAAACCAGTGTCATGAGTTGAATTC 10
|||||

RESULT 8
AX018075 359 bp DNA PAT 07-SEP-2000
LOCUS
DEFINITION Sequence 23 from Patent WO9946374.
ACCESSION AX018075
VERSION AX018075.1 GI:10042526
KEYWORDS
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 359)
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Pilarczyk,C.
TITLE Human nucleic acid sequences from prostate tumour tissue
JOURNAL Patent: WO 9946374-A 23 16-SEP-1999;
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILARSKY CHRISTIAN (DE)
FEATURES
Location/Qualifiers
1..359
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 121 a 75 c 94 g 69 t
ORIGIN
Query Match 6.6%; Score 235.4; DB 9; Length 359;
Best Local Similarity 99.6%; Pred. No. 6.7e-45;
Matches 236; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 281 cagggtgagaataagaagctgtgacttaccatctgagggccacacatctgctgaa 340
|||||
Db 123 CAGAGGTGAGAAATAAGAAAGCTGCTGACTTTACCATCTGAGGCCACACATCTGCTGAA 182
|||||
QY 341 atggagataataacatcactcagaacagcagatgacaaataataatgtctaagtgtgac 400
|||||
Db 183 ATGGAGATAATTAACATCACTACTAGAACACAGATGACAAATAATGTCTAAGTAGTGAC 242
|||||
QY 401 atgttttcacatttccagccctttaaataatccacacacagagaagcacaaaaggaa 460
|||||
Db 243 ATGTTTTTCACATTTCACAGCCCTTTAAATAATCCACACACAGGACCAAAAGGAA 302
|||||
QY 461 gcacagagatccctgggagaaatgccggccgcatcttgggtcatcgatgagcctc 517
|||||
Db 303 GCACAGAGATCCCTGGGAGAAATGCCGGCCGCGCATCTTTGGGTCTCATGATGAGCCCTC 359
|||||

RESULT 9
AL161625 143675 bp DNA HTG 20-JAN-2001
LOCUS
DEFINITION Homo sapiens chromosome 9 clone RP11-146P9, *** SEQUENCING IN
PROGRESS ***, 10 unordered pieces.
ACCESSION AL161625
VERSION AL161625.6 GI:9863607
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 143675)
AUTHORS Plumb,B.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerquest@sanger.ac.uk
COMMENT On Aug 21, 2000 this sequence version replaced gi:8894260.
----- Genome Center

Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA146P9
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 138647 bases at least Q40
Consensus quality: 140619 bases at least Q30
Consensus quality: 141706 bases at least Q20
Insert size: 142775; sum-of-contigs
Insert size: 147523; 5.9% error; agarose-fp
Quality coverage: 4.29x in Q20 bases; sum-of-contigs Quality
coverage: 4.22x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 12163: contig of 12163 bp in length
12164 12263: gap of 100 bp
12264 25269: contig of 13006 bp in length
25270 25369: gap of 100 bp
25370 57709: contig of 32340 bp in length
57710 57809: gap of 100 bp
57810 66792: contig of 8983 bp in length
66793 66892: gap of 100 bp
66893 102772: contig of 35880 bp in length
102773 102872: gap of 100 bp
102873 106863: contig of 3991 bp in length
106864 106963: gap of 100 bp
106964 129804: contig of 22841 bp in length
129805 129904: gap of 100 bp
129905 132223: contig of 2319 bp in length
132224 132323: gap of 100 bp
132324 135764: contig of 3441 bp in length
135765 135864: gap of 100 bp
135865 143675: contig of 7811 bp in length.

FEATURES
Source
1..143675
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-146P9"
/clone_lib="RPC1-11.1"
1..12163
/note="assembly_fragment:00440
clone_end:T7
vector_side:left"
12264..25269
/note="assembly_fragment:00434
fragment_chain:1"
25370..57709
/note="assembly_fragment:00526
fragment_chain:1"
57810..66792
/note="assembly_fragment:00661
fragment_chain:1"
66893..102772
/note="assembly_fragment:00323"
102873..106863
/note="assembly_fragment:00860"
106964..129804
/note="assembly_fragment:00978"
129905..132223
/note="assembly_fragment:01068"

misc_feature
1..12163
/note="assembly_fragment:00440
clone_end:T7
vector_side:left"
12264..25269
/note="assembly_fragment:00434
fragment_chain:1"
25370..57709
/note="assembly_fragment:00526
fragment_chain:1"
57810..66792
/note="assembly_fragment:00661
fragment_chain:1"
66893..102772
/note="assembly_fragment:00323"
102873..106863
/note="assembly_fragment:00860"
106964..129804
/note="assembly_fragment:00978"
129905..132223
/note="assembly_fragment:01068"

* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence,
* as soon as it is available and the accession number will
* be preserved.

FEATURES
source

Location/Qualifiers
1..267581
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-133022"
/clone_lib="RPCI-II.1"
1..25718
/note="assembly_fragment:00645
fragment_chain:1
clone_end:77
vector_side:left"
25819..29567
/note="assembly_fragment:01660
fragment_chain:1"
29668..40520
/note="assembly_fragment:03080
fragment_chain:1"
40621..50002
/note="assembly_fragment:00771
fragment_chain:1"
50103..54942
/note="assembly_fragment:02629
fragment_chain:1"
55043..59067
/note="assembly_fragment:04565
fragment_chain:1"
59168..72327
/note="assembly_fragment:00223
fragment_chain:1"
72428..79396
/note="assembly_fragment:03318
fragment_chain:1"
79497..82738
/note="assembly_fragment:00224
fragment_chain:2"
82839..87543
/note="assembly_fragment:03050
fragment_chain:2"
87644..89847
/note="assembly_fragment:01948
fragment_chain:3"
89948..92376
/note="assembly_fragment:02191
fragment_chain:3"
92477..99799
/note="assembly_fragment:00010"
99900..102117
/note="assembly_fragment:00672"
102218..109330
/note="assembly_fragment:01326"
109431..112187
/note="assembly_fragment:01498"
112288..116407
/note="assembly_fragment:01607"
116508..119386
/note="assembly_fragment:01899"
119487..121869
/note="assembly_fragment:02068"
121970..125546
/note="assembly_fragment:02469"
125647..128990
/note="assembly_fragment:02502"
129091..134055
/note="assembly_fragment:02669"
134156..138314
/note="assembly_fragment:02757"
138415..140612
/note="assembly_fragment:02776"
140713..143296
/note="assembly_fragment:03127"
143397..148113
/note="assembly_fragment:03634"

misc_feature 148214..151009
/note="assembly_fragment:03898"
misc_feature 151110..156616
/note="assembly_fragment:04220"
misc_feature 156717..158763
/note="assembly_fragment:04276"
misc_feature 158864..162129
/note="assembly_fragment:04418"
misc_feature 162230..164443
/note="assembly_fragment:04428"

Query Match 3.4%; Score 121; DB 80; Length 267581;
Best Local Similarity 100.0%; Pred. No. 7.2e-18;
Matches 121; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acagaagaatagcaagtgcgcgagagctgcacagaaaaacagagggagatttgt 60
|||||
Db 253253 ACAGAAGAATAAGCAAGTCCCGAGAGCTGGCATCAGAAAACAGAGGGAGATTGTGT 253312
|||||
QY 61 ggctgcagccgagggagaccaggaagatctgcattggtggaggacctgatatacagag 120
|||||
Db 253313 GCCTGCAGCCGAGGAGACCAGGAGATCTGCATGTTGGTGAAGGACCTGATCATACAGAG 253372
|||||
QY 121 g 121
Db 253373 G 253373

RESULT 11
AF279290 580 bp DNA PRI 04-DEC-2000
LOCUS Homo sapiens prostate-cancer-specific DD3 protein gene, promoter
DEFINITION region and exon 1.
ACCESSION AF279290
VERSION AF279290.1 GI:11528086
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 580)
AUTHORS Verhaegh,G.W., van Bokhoven,A., Smit,F., Schalken,J.A. and
Bussemakers,M.J.G.
TITLE Isolation and Characterization of the Promoter of the Human
Prostate Cancer-specific DD3 Gene
J. Biol. Chem. 275 (48), 37496-37503 (2000)
JOURNAL
PUBMED 10982808
REFERENCE 2 (bases 1 to 580)
AUTHORS Verhaegh,G.W., van Bokhoven,A., Smit,F., Schalken,J.A. and
Bussemakers,M.J.G.
TITLE Direct Submission
JOURNAL Submitted (16-JUN-2000) Urology Research Laboratory, University
Medical Center Nijmegen, Geert Grooteplein Zuid 10, Nijmegen 6525
GA, The Netherlands
FEATURES Location/Qualifiers
source
1..580
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/map="9q21-q22"
1..460
/bound_moiety="high mobility group I Y protein"
461..5580
/product="prostate-cancer-specific DD3 protein"
461..580
/number=1
BASE COUNT 190 a 97 c 153 g 140 t
ORIGIN

Query Match 3.4%; Score 120; DB 89; Length 580;
Best Local Similarity 100.0%; Pred. No. 1.1e-17;

Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acagagaataagcaagtccgagaaagctgcacgagaaacacagagggagattgtgt 60
|||||
Db 461 ACAGAGAATAAGCAAGTCCGAGAAAGCTGCATGTTGGGAAGGACCTGATGATACAGAG 520
|||||

QY 61 ggctgcagccgagggagaccagagaatctgcattggtgggaagacctgatgatacagag 120
|||||
Db 521 GGCTGCAGCCGAGGAGACCAAGGAAGATCTGCATGTTGGGAAGGACCTGATGATACAGAG 580
|||||

RESULT 12

AC009556/c

LOCUS

DEFINITION

AC009556

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC009556 172298 bp DNA HTG 26-MAY-2000
Homo sapiens clone RP11-57C21, WORKING DRAFT SEQUENCE, 26 unordered
pieces.
AC009556
HTG; HTGS_PHASE1; HTGS_DRAFT.
human.

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 172298)
Birren, B., Linton, L., Nusbaum, C., and Lander, E.
Homo sapiens chromosome, clone RP11-57C21
Unpublished

2 (bases 1 to 172298)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Brown, A.,
Castle, A., Cerny, J., Collangelo, M., Collins, S., Collymore, A.,
Cooke, P., DeArelano, K., Depayre, E., Devon, K., Dewar, K.,
Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C.,
Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G.,
Hagos, B., Heaford, A., Horton, L., Howland, J. C., Jones, C., Kann, L.,
Karatsis, A., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P.,
Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J.,
Meldrim, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,
Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,
Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,
Tesfaye, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,
Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.
Direct Submission

TITLE

JOURNAL

COMMENT

Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 28, 2000 this sequence version replaced gi:6479158.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1672
Center clone name: 57C.21
----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator; 5% of reads
Chemistry: Dye-terminator Big Dye; 95% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 136385 bases at least Q40
Consensus quality: 156543 bases at least Q30
Consensus quality: 165534 bases at least Q20
Insert size: 165000; agarose-1p
Insert size: 169798; sum-of-contents
Quality coverage: 4.2 in Q20 bases; agarose-1p
Quality coverage: 4.1 in Q20 bases.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
1016: contig of 1016 bp in length
1017 1116: gap of 100 bp
1117 2136: contig of 1020 bp in length
2137 2236: gap of 100 bp
2237 3571: contig of 1335 bp in length
3572 3671: gap of 100 bp
3672 5272: contig of 1601 bp in length
5273 5372: gap of 100 bp
5373 6381: contig of 1009 bp in length
6382 6481: gap of 100 bp
6482 8508: contig of 2027 bp in length
8509 8608: gap of 100 bp
8609 9886: contig of 1278 bp in length
9887 9986: gap of 100 bp
9987 11943: contig of 1957 bp in length
11944 12043: gap of 100 bp
12044 15112: contig of 3069 bp in length
15113 15212: gap of 100 bp
15213 19323: contig of 4111 bp in length
19324 19423: gap of 100 bp
19424 24065: contig of 4642 bp in length
24066 24165: gap of 100 bp
24166 31658: contig of 7493 bp in length
31659 31758: gap of 100 bp
31759 38737: contig of 6979 bp in length
38738 38837: gap of 100 bp
38838 46131: contig of 7294 bp in length
46132 46231: gap of 100 bp
46232 52344: contig of 6113 bp in length
52345 52444: gap of 100 bp
52445 59763: contig of 7319 bp in length
59764 59863: gap of 100 bp
59864 66816: contig of 6953 bp in length
66817 66916: gap of 100 bp
66917 74010: contig of 7094 bp in length
74011 74110: gap of 100 bp
74111 80599: contig of 6489 bp in length
80600 80699: gap of 100 bp
80700 89507: contig of 8808 bp in length
89508 89607: gap of 100 bp
89608 100337: contig of 10730 bp in length
100338 100437: gap of 100 bp
100438 114298: contig of 13861 bp in length
114299 114398: gap of 100 bp
114399 128530: contig of 14132 bp in length
128531 128630: gap of 100 bp
128631 142917: contig of 14287 bp in length
142918 143017: gap of 100 bp
143018 156243: contig of 13226 bp in length
156244 156343: gap of 100 bp
156344 172298: contig of 15955 bp in length.

FEATURES

source

1. 172298
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-57C21"
/clone_lib="RPCI-11 Human Male BAC"
misc_feature
1. 1016
/note="assembly_fragment"
1117. 2136
/note="assembly_fragment"
2237. 3571
/note="assembly_fragment"
3672. 5272
/note="assembly_fragment"
5373. 6381
/note="assembly_fragment"
6482. 8508
misc_feature

misc_feature /note="assembly_fragment"
8609. .9886
/note="assembly_fragment"
9987. .11943
/note="assembly_fragment"
12044. .15112
/note="assembly_fragment"
15213. .19323
/note="assembly_fragment"
19424. .24065
/note="assembly_fragment"
24166. .31658
/note="assembly_fragment"
31759. .38737
/note="assembly_fragment"
38838. .46131
/note="assembly_fragment"
46232. .52344
/note="assembly_fragment"
52445. .59763
/note="assembly_fragment"
59864. .66816
/note="assembly_fragment"
66917. .74010
/note="assembly_fragment"
clone_end:T7
vector_side:left
74111. .80599
/note="assembly_fragment"
clone_end:SP6
vector_side:left
80700. .89507
/note="assembly_fragment"
89608. .100337
/note="assembly_fragment"
100438. .114298
/note="assembly_fragment"
114399. .128530
/note="assembly_fragment"
128631. .142917
/note="assembly_fragment"
143018. .156243
/note="assembly_fragment"
156344. .172298
/note="assembly_fragment"
BASE COUNT 51330 a 34914 c 34487 g 49065 t 2502 others
ORIGIN

Query Match 3.3%; Score 119.4; DB 61; Length 172298;
Best Local Similarity 99.2%; Pred. No. 1.7e-17;
Matches 120; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 acagaagaatagcaagtccagagctgcatcagacagagggagatttgtgt 60
Db 16561 ACAGAAGAAATAGCAAGTCCGAGAGCTGGCATCAGAAAAACAGAGGGGAGATTGTGT 16502
Qy 61 ggcgcagcagagagaccaggaagatctgcatggtggagagacctgatgacagag 120
Db 16501 GGCCTGACCGGAGGAGACCGAGAGATTTCATGCTGGGAGGACCTGATGATACAGAG 16442
Qy 121 g 121
Db 16441 G 16441
RESULT 13
LOCUS I66494 7218 bp DNA PAT 28-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION I66494
VERSION I66494.1 GI:2724471
KEYWORDS
SOURCE
ORGANISM
Unclassified.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dorner,F., Scheifflinger,F. and Falkner,F.Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES Location/Qualifiers
source 1. .7218
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN

SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dorner,F., Scheifflinger,F. and Falkner,F.Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES Location/Qualifiers
source 1. .7218
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN
Query Match 2.2%; Score 77.4; DB 10; Length 7218;
Best Local Similarity 2.3%; Pred. No. 1.3e-07;
Matches 9; Conservative 245; Mismatches 131; Indels 0; Gaps 0;
Qy 2830 gaattctctattatgtagctgtctgtgcatctctgctctctctctcttgacacatat 2889
Db 1064 GATYVV 1123
Qy 2890 tagcttctagcctttgttccagcactttatctttctccacacatcgcttaccacatc 2949
Db 1124 YVV 1183
Qy 2950 ctctctctgtctgtgttggacttcccccacagaatttcaacgactctcaagtcttt 3009
Db 1184 YVV 1243
Qy 3010 tottccatcccccacacataactgaatgcctagacccttatttttaattccacata 3069
Db 1244 YVV 1303
Qy 3070 gatgctgctatgggctaataattgcttagatgaacattagatatattaaagctcaagagg 3129
Db 1304 YVV 1363
Qy 3130 tcaaaatccaaactattatctctctctctctctctctctctctctctctctctctct 3189
Db 1364 YVV 1423
Qy 3190 actgattgactgaacagagatggtcc 3214
Db 1424 YVVVVVVVVVGTACCAAAATCTTC 1448
RESULT 14
LOCUS I66494/c 7218 bp DNA PAT 28-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION I66494
VERSION I66494.1 GI:2724471
KEYWORDS
SOURCE
ORGANISM
Unclassified.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dorner,F., Scheifflinger,F. and Falkner,F.Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES Location/Qualifiers
source 1. .7218
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN

Query Match 1.6%; Score 57.4; DB 10; Length 7218;
Best Local Similarity 4.3%; Pred. No. 0.0066;
Matches 16; Conservative 211; Mismatches 142; Indels 0; Gaps 0;
Qy 444 aggaagcacaagaagcacagagatccctgggagaaatgccgcgccttcttgggt 503

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 16:53:08 ; Search time 46.06 Seconds
(without alignments)
146.495 Million cell updates/sec

Title: US-09-402-713a-7
Perfect score: 268
Sequence: 1 MFLHSSPKYKPYHTQBAQKE.....HLGSSMSLALCLVPLVREGH 51

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL16:*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp-invertebrate:*
6: sp-mammal:*
7: sp-mhc:*
8: sp-organelle:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-unclassified:*
13: sp-vertebrate:*
14: sp-virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	60.5	22.6	1236	5 Q9NKF9	Q9NKF9 drosophila
2	60.5	22.6	1238	5 Q9VJW9	Q9VJW9 drosophila
3	60.5	22.6	1239	5 Q94902	Q94902 drosophila
4	60	22.4	187	2 Q9JUF3	Q9JUF3 neisseria m
5	60	22.4	218	2 Q9JRV5	Q9JRV5 neisseria m
6	59	22.0	1114	11 Q9JL73	Q9JL73 mus musculus
7	56.5	21.1	1061	2 P73998	P73998 synecocyst
8	56.5	21.1	6396	2 Q9KID7	Q9KID7 streptocyst
9	56	20.9	338	2 Q9RVU4	Q9RVU4 deinococcus
10	56	20.9	433	5 Q9V7N7	Q9V7N7 drosophila
11	56	20.9	598	5 Q9NFPI	Q9NFPI drosophila
12	55.5	20.7	658	2 Q9F1J5	Q9F1J5 pseudomonas
13	55	20.5	168	14 Q9E138	Q9E138 bovine vira
14	55	20.5	196	5 Q9VKF5	Q9VKF5 drosophila
15	55	20.5	469	5 Q9GR75	Q9GR75 leishmania
16	55	20.5	1032	11 Q61989	Q61989 mus musculus
17	55	20.5	1935	5 Q9VQW0	Q9VQW0 drosophila
18	54.5	20.3	116	14 Q92886	Q92886 human immun
19	54.5	20.3	177	6 Q9XS95	Q9XS95 oryctolagus

20	54.5	20.3	244	2 Q9HYW1	Q9HYW1 pseudomonas
21	54.5	20.3	305	10 Q9M1C3	Q9M1C3 arabisopsis
22	54.5	20.3	461	11 Q9R1Y8	Q9R1Y8 mus musculus
23	54	20.1	204	3 Q9P3L8	Q9P3L8 neurospora
24	54	20.1	257	2 Q9Z8E3	Q9Z8E3 chlamydia p
25	54	20.1	450	10 Q9FMN6	Q9FMN6 pseudomonas
26	54	20.1	633	2 Q9HTU6	Q9HTU6 pseudomonas
27	54	20.1	784	4 Q00146	Q00146 homo sapien
28	53.5	20.0	250	8 Q33122	Q33122 schismocarp
29	53.5	20.0	348	8 Q35379	Q35379 paracyprieh
30	53.5	20.0	408	4 Q9UGL4	Q9UGL4 homo sapien
31	53.5	20.0	421	8 Q9TIC2	Q9TIC2 balbiana i
32	53.5	20.0	443	8 Q32878	Q32878 palmaria pa
33	53.5	20.0	449	2 Q9Z9E0	Q9Z9E0 chlamydia p
34	53.5	20.0	783	4 Q9NXD0	Q9NXD0 homo sapien
35	53	19.8	671	5 Q9W365	Q9W365 drosophila
36	53	19.8	1161	11 Q9QYE7	Q9QYE7 rattus norv
37	53	19.8	2205	5 Q9NGQ2	Q9NGQ2 dictyosteli
38	52.5	19.6	216	2 Q9RVK0	Q9RVK0 deinococcus
39	52.5	19.6	357	2 Q9I720	Q9I720 pseudomonas
40	52.5	19.6	360	4 Q9H9N9	Q9H9N9 homo sapien
41	52.5	19.6	384	8 Q32401	Q32401 halosaccion
42	52.5	19.6	392	8 Q31855	Q31855 audouinella
43	52.5	19.6	424	8 Q9MV15	Q9MV15 ballia call
44	52.5	19.6	488	8 Q9THF8	Q9THF8 palmaria pa
45	52.5	19.6	493	4 Q9UF35	Q9UF35 homo sapien

ALIGNMENTS

RESULT 1
Q9NKF9 PRELIMINARY; PRT: 1236 AA.
ID Q9NKF9
AC Q9NKF9;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE KUZ PROTEIN.
GN KUZ OR BG:DS07660.3 OR CG7147.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y, CN BW SP;
RX MEDLINE=99403001; PubMed=10471707;
RA Ashburner M., Misra S., Roote J., Lewis S.E., Blazej R., Davis T.,
RA Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,
RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
RA Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitelaw K.,
RA Ceiniker S., Rubin G.M.;
RT "An exploration of the sequence of a 2.9-Mb region of the genome of
RT Drosophila melanogaster: the Adh region.";
RL Genetics 153:179-219(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Y, CN BW SP;
RA Ceiniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,
RA Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
RA Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
RA Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,
RA Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
RA Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,
RA Sethi H., Snir E., Svirska R.R., Wan K.H., Weinburg T., Zhang R.,
RA Zieran L.L., Rubin G.M.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003407; AAF44800.1;
DR Flybase; FBgn0015954; kuz.
DR InterPro; IPR000130;
DR InterPro; IPR001590;

```
DR InterPro; IPR001762; -.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS50214; DISINTEGRIN_2; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR SMART; SM00050; DISIN; 1.
SQ SEQUENCE 1236 AA; 136343 MW; 40461AC17040C9AD CRC64;

Query Match      22.6%; Score 60.5; DB 5; Length 1236;
Best Local Similarity 47.2%; Pred. No. 19;
Matches 17; Conservative 4; Mismatches 10; Indels 5; Gaps 2;

QY 9 FKYPHTQEAQKQAQRSLGEM----PGRHLGSSMSLA 40
Db 338 FKYPH-QKYTKANFAEGAFYDPSTGRRLLGSSANVA 372
||||| I: ||| : | |||| :|

RESULT 2
QYVJW9 PRELIMINARY; PRT; 1238 AA.
AC Q9VJW9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE KUZ PROTEIN.
GN KUZ OR BG:DS07660.3 OR CG7147.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Abril J.F., Abmayyan A., An H.-J., Andrews-Pfannkoch C., Bakos G.L.G.,
RA Ballwey L.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler K., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaiswal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
```

```
DR EMBL; AE003640; AAF53318.1; -.
DR HSP; P17494; 1KST.
DR FlyBase; FBgn0015954; kuz.
DR InterPro; IPR000130; -.
DR InterPro; IPR001590; -.
DR InterPro; IPR001762; -.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS50214; DISINTEGRIN_2; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR SMART; SM00050; DISIN; 1.
SQ SEQUENCE 1238 AA; 136407 MW; 96D924B6188AC472 CRC64;

Query Match      22.6%; Score 60.5; DB 5; Length 1238;
Best Local Similarity 47.2%; Pred. No. 19;
Matches 17; Conservative 4; Mismatches 10; Indels 5; Gaps 2;

QY 9 FKYPHTQEAQKQAQRSLGEM----PGRHLGSSMSLA 40
Db 340 FKYPH-QKYTKANFAEGAFYDPSTGRRLLGSSANVA 374
||||| I: ||| : | |||| :|

RESULT 3
QY4902 PRELIMINARY; PRT; 1239 AA.
ID QY4902;
AC QY4902;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE KUZBANIAN.
GN KUZ OR BG:DS07660.3 OR CG7147.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96355647; PubMed=8703057;
RA Rooke J., Pan D., Xu T., Rubin G.M.;
RT "KUZ, a conserved metalloprotease-disintegrin protein with two roles
RT in Drosophila neurogenesis.";
RL Science 273:1227-1231(1996).
DR EMBL; U60591; AAC47275.1; -.
DR HSP; P17494; 1KST.
DR MEROPS; M12.211; -.
DR FlyBase; FBgn0015954; kuz.
DR InterPro; IPR000130; -.
DR InterPro; IPR001590; -.
DR InterPro; IPR001762; -.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS50214; DISINTEGRIN_2; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR SMART; SM00050; DISIN; 1.
KW Neurogenesis.
SQ SEQUENCE 1239 AA; 136522 MW; 7FCB514A1352577A CRC64;

Query Match      22.6%; Score 60.5; DB 5; Length 1239;
Best Local Similarity 47.2%; Pred. No. 19;
Matches 17; Conservative 4; Mismatches 10; Indels 5; Gaps 2;

QY 9 FKYPHTQEAQKQAQRSLGEM----PGRHLGSSMSLA 40
Db 341 FKYPH-QKYTKANFAEGAFYDPSTGRRLLGSSANVA 375
||||| I: ||| : | |||| :|

RESULT 4
QYJUF3 PRELIMINARY; PRT; 187 AA.
ID QYJUF3;
AC QYJUF3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
```



```

DR Pfam: PF00873; ACR_tran; 1.
DR PROSITE; PS00702; ACRIFLAVINRP.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
SQ SEQUENCE 1061 AA; 114935 MW; 480D9B60FC09F8E2 CRC64;

Query Match 21.18; Score 56.5; DB 2; Length 1061;
Best Local Similarity 31.08; Pred. No. 55;
Matches 13; Conservative 11; Mismatches 17; Indels 1; Gaps 1;

QY 5 ISSPFKYPHTQEAQKAEQRSIGEMPGRHGLSSMSLALCLVPL 46
   || | | | | | | | | | | | | | | | | | | | |
Db 418 ISRFITKVKH-EDPQEAQEAANGELTGAVIATSLVLMVFI 458

RESULT 8
Q9KID7 PRELIMINARY; PRT; 6396 AA.
ID Q9KID7 AC
Q9KID7 AC
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE FREA.
GN FREA.
OS Streptomyces hygroscopicus var. ascomyceticus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=132248;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20323220; PubMed=10863099;
RA Wu K., Chung L., Revilli W.P., Katz L., Reeves C.D.;
RT "The FK520 gene cluster of streptomyces hygroscopicus var.
RT ascomyceticus (ATCC 14891) contains genes for biosynthesis of unusual
RT polyketide extender units.";
RL Gene 251:81-90(2000).
DR EMBL; AF235504; AAF86396.1; -.
DR InterPro; IPR000255; -.
DR InterPro; IPR000794; -.
DR InterPro; IPR001066; -.
DR InterPro; IPR001227; -.
DR InterPro; IPR001899; -.
DR InterPro; IPR002085; -.
DR Pfam; PF00107; adh_zinc; 2.
DR Pfam; PF00109; ketoacyl-synt; 4.
DR Pfam; PF00550; pp-binding; 4.
DR Pfam; PF00698; Acyl_transf; 4.
DR PROSITE; PS00075; ACP_DOMAIN; 4.
DR PROSITE; PS00608; B_KETOACYL_SYNTHASE; 4.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_3.
DR PROSITE; PS00012; PHOSPHANTHETHEINE; UNKNOWN_3.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW Phosphoantethine; Transferase.
SQ SEQUENCE 6396 AA; 669881 MW; FF13BC2431D2E805 CRC64;

Query Match 21.18; Score 56.5; DB 2; Length 6396;
Best Local Similarity 35.68; Pred. No. 3.9e+02;
Matches 16; Conservative 6; Mismatches 14; Indels 9; Gaps 2;

QY 12 PHT-----QEAQKAEQRSIG---EMPGRHGLSSMSLALCLVPL 47
   || | | | | | | | | | | | | | | | | | | | |
Db 653 PHSVLVSGDEAVLEAARQLGIHHRLPTRHAGSERMQPLVAPLL 697

RESULT 9
Q9RVU4 PRELIMINARY; PRT; 338 AA.
ID Q9RVU4 AC
Q9RVU4 AC
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE ENDONUCLEASE III, PUTATIVE.

```

```

Query Match      20.9%; Score 56; DB 5; Length 558;
Best Local Similarity 33.3%; Pred. No. 32;
Matches 12; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

QY 3 LHJSSPFKYPHPTQPAQKEAQRSLGEMPCRHJLGSSMS 38
   :::  ||  ::  :||| ||  :  ||  |
Db 292 IYVAQVPKYTRSLKKRAKRSQEIOTROVASSSS 327

```

```

Query Match      20.5%; Score 55; DB 14; Length 168;
Best Local Similarity 40.5%; Pred. No. 12;
Matches 15; Conservative 4; Mismatches 14; Indels 4; Gaps 2;

Qy 3 LHITSPFKYPHTQBAKQAQRSLGEMPR---HLGSS 36
   | | | | | | | | | | | | | | | | | |
Db ~ 39 LHQSTLKLPH-KRGEKEVPTSLASLPKRGDCRLGNS 74

RESULT 14
Q9VKKF5
ID Q9VKF5 PRELIMINARY; PRT; 196 AA.
AC Q9VKF5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)

```

[illegible]

Search completed: July 30, 2001, 16:53:08
Job time: 188 sec

Query Match 20.5%; Score 55; DB 5; Length 196;
Best Local Similarity 61.1%; Pred. No. 14;
Matches 11; Conservative 4; Mismatches 3; Indels

RESULT 15
Q9GRT5
PRELIMINARY;
PRT; 469 AA.

Q5GRJ5, 01-MAR-2001 (TREMBLrel. 16, Created)
 DT DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE CYCLIN-DEPENDENT KINASE.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM:protein - protein search, using sw model

Run on: July 30, 2001, 16:53:31 ; Search time 17.83 Seconds
(without alignments)
97.983 Million cell updates/sec

Title: US-09-402-713A-7

Perfect score: 268

Sequence: 1 MFLHISPPFKYPTQEAQKE.....HLGSSMSLALCLVPLVREGH 51

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	60.5	22.6	343	1	TA2R_BOVIN
2	59	22.0	1113	1	HDAS_MOUSE
3	59	22.0	1122	1	HDAS_HUMAN
4	55	20.5	1039	1	ITR4_MOUSE
5	54.5	20.3	461	1	COLA_BOVIN
6	54.5	20.3	461	1	COLA_HUMAN
7	54.5	20.3	461	1	COLA_MOUSE
8	53.5	20.0	446	1	GNLU_ECOLI
9	53.5	20.0	451	1	PTCC_BACST
10	53.5	20.0	580	1	5NTD_BOOMI
11	53	19.8	554	1	DCS4_GOSAR
12	52.5	19.6	441	1	RXRA_BRARE
13	52	19.4	337	1	RIR2_TRYBB
14	52	19.4	2774	1	MAPA_RAT
15	51.5	19.2	269	1	AQPL_HUMAN
16	51.5	19.2	470	1	RBL_PROHO
17	51.5	19.2	569	1	4CL2_ORISA
18	51	19.0	352	1	KE4_BRARE
19	51	19.0	1038	1	ITR4_HUMAN
20	51	19.0	1162	1	ITR4_HUMAN
21	51	19.0	2485	1	POLN_EEVV3
22	51	19.0	2492	1	POLN_EEVVP
23	50.5	18.8	409	1	AXIA_BRARE
24	50.5	18.8	554	1	DCS2_GOSAR
25	50.5	18.8	2472	1	NCR2_MOUSE
26	50	18.7	130	1	ITP_SCHGR
27	50	18.7	134	1	ITPL_SCHGR
28	50	18.7	456	1	HMFT_DROHY
29	50	18.7	467	1	RBL_PHORE
30	50	18.7	528	1	ANM3_RAT
31	50	18.7	1100	1	JAK3_RAT
32	50	18.7	1163	1	ITAX_HUMAN
33	50	18.7	2492	1	POLN_EEVVT

34	49.5	18.5	490	1	CPCC_RAT	P11510	rattus norv
35	49.5	18.5	499	1	RPB2_METVA	P41558	methanococc
36	49.5	18.5	554	1	DCS1_GOSHI	P93665	gossypium h
37	49.5	18.5	741	1	IDH2_VIBAI	P41561	vibrio sp.
38	49.5	18.5	893	1	PER_PERAM	Q25637	periplaneta
39	49	18.3	221	1	SCAB_SHEEP	O62816	ovis aries
40	49	18.3	405	1	PEZ1_RAT	P70597	rattus norv
41	49	18.3	608	1	GLMS_ECOLI	P17169	e glucosami
42	48.5	18.1	548	1	ILVG_ECOLI	P00892	escherichia
43	48.5	18.1	554	1	DCS1_GOSAR	Q39761	gossypium a
44	48.5	18.1	644	1	RNR_HELPJ	O921x9	helicobacte
45	48.5	18.1	834	1	CASL_HUMAN	Q14511	homo sapien

ALIGNMENTS

RESULT 1

TA2R_BOVIN

ID TA2R_BOVIN STANDARD; PRT; 343 AA.

AC Q95125; Q95124;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE THROMBOXANE A2 RECEPTOR (TXA2-R) (PROSTANOID TP RECEPTOR).

GN TBXA2R.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Bovidae; Bovinae; Bos.

OX NCBI_TaxID=9913;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Liver, and Heart;

RX MEDLINE=98221872; PubMed=95611101;

RA Muck S., Schroer K.;

RT "Cloning, tissue-specific expression and regulation of the bovine thromboxane A2 receptor";

RL Adv. Exp. Med. Biol. 433:47-50(1997).

CC -!- FUNCTION: RECEPTOR FOR THROMBOXANE A2 (TXA2), A POTENT STIMULATOR OF PLATELET AGGREGATION. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY A G-PROTEIN THAT ACTIVATE A PHOSPHATIDYLINOSITOL-CALCIUM SECOND MESSENGER SYSTEM. IN THE KIDNEY, THE BINDING OF TXA2 TO GLOMERULAR TP RECEPTORS CAUSES INTENSE VASOCONSTRICTION.

CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; U53485; AAC34309.1; -

CC EMBL; U53484; AAC34308.1; -

CC GCRdb; GCR.1217; -

DR	InterPro; IPR000276; -						
DR	InterPro; IPR001105; -						
DR	Pfam; PF00001; 7tm1; 1.						
DR	PRINTS; PRO0429; THROMBOXANER.						
DR	PROSITE; PS00237; G-PROTEIN_RECEP_F1_1; 1.						
DR	PROSITE; PS00262; G-PROTEIN_RECEP_F1_2; 1.						
KW	G-protein coupled receptor; Transmembrane; Glycoprotein.						
FT	DOMAIN 1 29						
FT	TRANSMEM 30 52						
FT	DOMAIN 53 66						
FT	DOMAIN 67 87						
FT	TRANSMEM 88 106						
FT	DOMAIN 107 128						
FT	TRANSMEM 129 149						
FT	DOMAIN 150 172						
FT	TRANSMEM 172						

GN CORO1A OR CORO1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Spleen;
RX MEDLINE=95278344; PubMed=7758584;
RA Suzuki K., Nishihata J., Arai Y., Honma N., Yamamoto K., Irimura T.,
RA Toyoshima S.;
RT "Molecular cloning of a novel actin-binding protein, p57, with a WD
RT repeat and a leucine zipper motif.";
RL FEBS Lett. 364:283-288(1995).
CC -1- FUNCTION: MAY PLAY A ROLE IN THE SIGNAL TRANSDUCTION PATHWAYS OF
CC CHEMOTAXIS.
CC -1- SUBUNIT: BINDS ACTIN.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, THYMUS, SPLEEN, BONE
CC MARROW AND LYMPH NODE. LOW IN LUNG AND GUT.
CC -1- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
CC -1- SIMILARITY: BELONGS TO THE CORONIN FAMILY OF WD-REPEAT PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D44496; BAA07939.1; -
DR InterPro; IPR001680; -
DR Pfam; PF00400; WD40; 3.
DR PROSITE; PS00678; WD_REPEATS_1; 2.
DR PROSITE; PS50082; WD_REPEATS_2; 2.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Actin-binding; Repeat; WD repeat; Coiled coil.
FT REPEAT 79 119 WD 1.
FT REPEAT 129 169 WD 2.
FT REPEAT 174 213 WD 3.
FT REPEAT 218 260 WD 4.
FT REPEAT 265 305 WD 5.
FT DOMAIN 424 460 COILED COIL (POTENTIAL).
FT SEQUENCE 461 AA; 50979 MW; 1721A5D093C1130A CRC64;
Query Match 20.3%; Score 54.5; DB 1; Length 461;
Best Local Similarity 43.8%; Pred. No. 13;
Matches 14; Conservative 3; Mismatches 8; Indels 7; Gaps 1;
QY 2 FLHISPPKYPHTQEAQKEAQRSLGEMPGRHL 33
DB 303 FLHYSMF-----SSKESQRCGMGYMPKRG 327
RESULT 6
CO1A_HUMAN STANDARD; PRT; 461 AA.
AC P31146;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CORONIN-LIKE PROTEIN P57 (CORONIN 1A).
GN CORO1A OR CORO1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Periphereal blood leukocytes;
RX MEDLINE=95278344; PubMed=7758584;

RA Suzuki K., Nishihata J., Arai Y., Honma N., Yamamoto K., Irimura T.,
RA Toyoshima S.;
RT "Molecular cloning of a novel actin-binding protein, p57, with a WD
RT repeat and a leucine zipper motif.";
RL FEBS Lett. 364:283-288(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX Grodan A., Keep N.H., Reeves E., Segal A.W.;
RA Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Liau G., Popa I., Argraves K., Argraves W.S.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 355-374.
RC TISSUE=Keratinocytes;
RX Rasmussen H.H., van Damme J., Puype M., Gesser B., Celis J.E.,
RA Vandekerckhove J.;
RT "Microsequences of 145 proteins recorded in the two-dimensional gel
RT protein database of normal human epidermal keratinocytes.";
RL Electrophoresis 13:960-969(1992).
CC -1- FUNCTION: MAY PLAY A ROLE IN THE SIGNAL TRANSDUCTION PATHWAYS OF
CC CHEMOTAXIS (BY SIMILARITY).
CC -1- SUBUNIT: BINDS ACTIN (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, THYMUS, SPLEEN, BONE
CC MARROW AND LYMPH NODE. LOW IN LUNG AND GUT.
CC -1- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
CC -1- SIMILARITY: BELONGS TO THE CORONIN FAMILY OF WD-REPEAT PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D44497; BAA07940.1; -
DR EMBL; X89109; CAA61482.1; -
DR EMBL; U34690; AAA77058.1; -
DR MIM; 605000; -
DR Aarhus/Ghent-2DPAGE; 2416; IEF.
DR InterPro; IPR001680; -
DR Pfam; PF00400; WD40; 3.
DR PROSITE; PS00678; WD_REPEATS_1; 2.
DR PROSITE; PS50082; WD_REPEATS_2; 2.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Actin-binding; Repeat; WD repeat; Coiled coil.
FT REPEAT 79 119 WD 1.
FT REPEAT 129 169 WD 2.
FT REPEAT 174 213 WD 3.
FT REPEAT 218 260 WD 4.
FT REPEAT 265 305 WD 5.
FT DOMAIN 424 460 COILED COIL (POTENTIAL).
FT VARIANT 355 355 K -> D.
FT VARIANT /FTID=VAR_007722.
FT VARIANT 356 356 S -> D.
FT /FTID=VAR_007723.
FT CONFLICT 8 8 S -> T (IN REF. 3).
FT CONFLICT 245 245 R -> W (IN REF. 3).
SQ SEQUENCE 461 AA; 51026 MW; DE3FEDA57041515E CRC64;
Query Match 20.3%; Score 54.5; DB 1; Length 461;
Best Local Similarity 43.8%; Pred. No. 13;
Matches 14; Conservative 3; Mismatches 8; Indels 7; Gaps 1;
QY 2 FLHISPPKYPHTQEAQKEAQRSLGEMPGRHL 33
DB 303 FLHYSMF-----SSKESQRCGMGYMPKRG 327


```

RESULT 7
COLA_MOUSE
ID COLA_MOUSE STANDARD; PRT; 461 AA.
AC O89053; Q9R288;
DT 30-MAY-2000 (Rel. 39, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DE LOW-AFFINITY GLUCONATE TRANSPORTER (GLUCONATE PERMEASE) (GNT-I SYSTEM).
DE CORONIN-LIKE PROTEIN P57 (CORONIN 1A).
GN COROLA OR CORO1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98449467; PubMed=9778037;
RA Okumura M., Kung C., Wong S., Rodgers M., Thomas M.L.;
RT "Definition of family of coronin-related proteins conserved between humans and mice: close genetic linkage between coronin-2 and CD45-associated protein."
RT associated protein."
RL DNA Cell Biol. 17:779-787(1998).
RN [2]
RP SEQUENCE OF 156-276 FROM N.A.
RC STRAIN=BALB/C; TISSUE=Spleen;
RX MEDLINE=99012997; PubMed=9798653;
RA Chu C.C., Paul W.E.;
RT "Expressed genes in interleukin-4 treated B cells identified by cDNA representation difference analysis."
RT Mol. Immunol. 35:487-502(1998).
RL Mol. Immunol. 35:487-502(1998).
CC -!- FUNCTION: MAY PLAY A ROLE IN THE SIGNAL TRANSDUCTION PATHWAYS OF CHEMOTAXIS (BY SIMILARITY).
CC -!- SURUNIT: BINDS ACTIN (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: EXPRESSION RESTRICTED TO HEMATOPOIETIC CELLS.
CC -!- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
CC -!- SIMILARITY: BELONGS TO THE CORONIN FAMILY OF WD-REPEAT PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF143955; AAD32703.1;
DR EMBL; U89399; AAC36506.1;
DR MGD; MGI:1345961; Corola.
DR InterPro; IPR001680;
DR Pfam; PF00400; WD40; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 2.
DR PROSITE; PS00082; WD_REPEATS_2; 2.
DR PROSITE; PS02994; WD_REPEATS_REGION; 1.
KW Actin-binding; Repeat; WD repeat; Coiled coil.
FT REPEAT 79 119 WD 1.
FT REPEAT 129 169 WD 2.
FT REPEAT 174 213 WD 3.
FT REPEAT 218 260 WD 4.
FT REPEAT 265 305 WD 5.
FT DOMAIN 424 460 COILED COIL (POTENTIAL).
FT SEQUENCE 461 AA; 50975 MW; 51D48ED91E6FE82A CRC64;
SQ

```

Query Match 20.3%; Score 54.5; DB 1; Length 461;

Best Local Similarity 43.8%; Pred. NO. 13;

Matches 14; Conservative 3; Mismatches 8; Indels 7; Gaps 1;

OY 2 FLHISPPKYPHTQBAQKAEQSLGEMPGRL 33

Db 303 FLHLSMF-----SSKESQRMGMVPRGL 327

```

RESULT 8
GNTU_ECOLI
ID GNTU_ECOLI STANDARD; PRT; 446 AA.
AC P46858; P76694; P76695;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE LOW-AFFINITY GLUCONATE TRANSPORTER (GLUCONATE PERMEASE) (GNT-I SYSTEM).
DE GNTU OR B3435/B3436.
GN Escherichia coli.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=97280784; PubMed=9135111;
RA Izu H., Adachi O., Yamada M.;
RT "Gene organization and transcriptional regulation of the gntRku operon involved in gluconate uptake and catabolism of Escherichia coli."
RT J. Mol. Biol. 267:778-793(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=96236044; PubMed=8655507;
RA Tong S., Porco A., Isturiz T., Conway T.;
RT "Cloning and molecular genetic characterization of the Escherichia coli gntR, gntK, and gntU genes of GntI, the main system for gluconate metabolism."
RL J. Bacteriol. 178:3260-3269(1996).
CC -!- FUNCTION: PART OF THE GLUCONATE UTILIZATION SYSTEM GNT-I; LOW-AFFINITY INTAKE OF GLUCONATE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE GNTP FAMILY OF PERMEASES.
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A FRAMESHIFT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D84362; BAA12326.1;
DR EMBL; U18997; AAA58234.1; ALT_FRAME.
DR EMBL; U18997; AAA58233.1; ALT_FRAME.
DR EMBL; AE000420; AAC76461.1; ALT_FRAME.
DR EMBL; AE000420; AAC76460.1; ALT_FRAME.
DR EcoGene; EGI2631; gntU.
KW Gluconate utilization; Sugar transport; Transport; Transmembrane; Inner membrane; Multigene family.
FT TRANSMEM 2 22 POTENTIAL.
FT TRANSMEM 27 47 POTENTIAL.
FT TRANSMEM 59 79 POTENTIAL.
FT TRANSMEM 110 130 POTENTIAL.
FT TRANSMEM 143 163 POTENTIAL.
FT TRANSMEM 177 197 POTENTIAL.
FT TRANSMEM 226 246 POTENTIAL.
FT TRANSMEM 262 282 POTENTIAL.

```

```

Query Match          20.0%: Score 53.5; DB 1; Length 451;
Best Local Similarity 28.0%: Pred. No. 17;
Matches 14; Conservative 12; Mismatches 19; Indels 5; Gaps 1;

QY 5 ISSPEKYPHTQEAQKQAGSLGEMPGRHLSG-----SMTSLCLVPLVRE 49
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 116 LATPYQVFTPEGAKETIMVSGGIPVQWVSGKGLFVAMILAIVSTEIYRK 165
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
5NTD_BOOMI          STANDARD;          PRT;          580 AA.
AC P52307; P90696;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE 5'-NUCLEOTIDASE PRECURSOR (EC 3.1.3.5) (FRAGMENT)..
OS Boophilus microplus (Cattle tick).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Parasitiformes; Ixodida; Ixodidae; Boophilus.
ON NCBI_TaxID=6941;
RX [1]
RX MEDLINE=99306777; PubMed=10380109;
RA Llyou N., Hamilton S., Elvin C., Willadsen P.;
RT "Cloning and expression of ecto 5-nucleotidase from the cattle tick
RL Boophilus microplus.";
RL Insect Mol. Biol. 8:257-266(1999).
RX [2]
RX MEDLINE=15-40 AND 162-180.
RX MEDLINE=93250870; PubMed=8387372;
RA Willadsen P., Riding G.A., Jarney J., Atkins A.;
RT "The nucleotidase of Boophilus microplus and its relationship to
RL enzymes from the rat and Escherichia coli.";
RL Insect Biochem. Mol. Biol. 23:291-295(1993).
CC -1- FUNCTION: DEGRADATION OF EXTERNAL UDP-GLUCOSE TO URTIDINE
CC MONOPHOSPHATE AND GLUCOSE-1-PHOSPHATE, WHICH CAN THEN BE USED BY
CC THE CELL.
CC -1- CATALYTIC ACTIVITY: UDP-SUGAR + H(2)O = UMP + SUGAR 1-PHOSPHATE.
CC -1- CATALYTIC ACTIVITY: A 5'-RIBONUCLEOTIDE + H(2)O = A
CC RIBONUCLEOSIDE + ORTHOPHOSPHATE.
CC -1- COFACTOR: ZINC.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -1- TISSUE SPECIFICITY: GUT, OVARIES AND SALIVARY GLANDS.
CC -1- PTM: GLYCOSYLATED.
CC -1- SIMILARITY: BELONGS TO THE 5'-NUCLEOTIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U80634; AAB38963.1; -.
CC InterPro; IPR000934; -.
CC InterPro; IPR002224; -.
CC PROSITE; PS00785; 5_NUCLEOTIDASE_1; FALSE_NEG.
CC PROSITE; PS00786; 5_NUCLEOTIDASE_2; FALSE_NEG.
CC Pfam; PF01009; 5_nucleotidase; 1.
CC Hydrolase; Gpi-anchor; Glycoprotein; Signal; Zinc.
CC NON_TER 1 1
CC FT SIGNAL 1 14
CC FT CHAIN 15 552
CC FT PROPEP 553 580
CC FT LIPID 552 552
CC FT CARBOHYD 172 172
CC FT CARBOHYD 285 285
CC FT CARBOHYD 423 423
CC FT CARBOHYD 536 536
CC
CC 5'-NUCLEOTIDASE.
CC REMOVED IN MATURE FORM (BY SIMILARITY).
CC GPI-ANCHOR (POTENTIAL).
CC N-LINKED (GLCNAC. .) (PROBABLE).
CC N-LINKED (GLCNAC. .) (POTENTIAL).
CC N-LINKED (GLCNAC. .) (POTENTIAL).
CC N-LINKED (GLCNAC. .) (POTENTIAL).

```

```
FT CONFLICT 15 15 T -> K (IN REF. 2).
FT CONFLICT 37 39 SGT -> HXG (IN REF. 2).
SQ SEQUENCE 580 AA; 63460 MW; 588EEF2014071AB7 CRC64;

Query Match 20.0%; Score 53.5; DB 1; Length 580;
Best Local Similarity 36.1%; Pred. No. 23;
Matches 22; Conservative 5; Mismatches 21; Indels 13; Gaps 3;

QY 4 HISSPFKYPHTQEA---KQEAORSIGE-----MGRHLGSSMSLAIC-----LVLVREG 50
Db 175 YLSSPGKVRFTDEACIOREAGRLREECVQVIIAGVSGVPRDLEICRVPVSLVGG 234

QY 51 H 51
Db 235 H 235

RESULT 11
DCS4_GOSAR STANDARD; PRT; 554 AA.
AC Q49853;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE (+)-DELTA-CADINENE SYNTHASE ISOZYME C2 (EC 4.6.1.11) (D-CADINENE
DE SYNTHASE).
GN CADI-C2
OS Gossypium arboreum.
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Malvales; Malvaceae; Gossypium.
OX NCBI_TaxID=29729;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NANKING;
RA Meng Y., Jia J., Liu C., Liang W., Zhou X., Heinstein P., Chen X.-Y.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: (E,E)-FARNESYL DIPHOSPHATE CYCLIZING.
CC -!- CATALYTIC ACTIVITY: 2-TRANS,6-TRANS-FARNESYL DIPHOSPHATE = (+)-
CC DELTA-CADINENE + DIPHOSPHATE.
CC -!- PATHWAY: FIRST COMMITTED STEP IN THE PATHWAYS LEADING TO THE
CC RELATED PHYTOALEXINS GOSSYPOL AND LACINILENE C.
CC -!- SIMILARITY: TO OTHER PLANT TERPENE CYCLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y16432; CAA76223.1; -.
DR HSSP; Q40577; 5EAD.
DR InterPro; IPR001906; -.
DR Pfam; PF01397; Terpene_synth; 1.
KW Lyase; Multigene family.
SQ SEQUENCE 554 AA; 64117 MW; 35DDDD66D3E838AAC CRC64;

Query Match 19.8%; Score 53; DB 1; Length 554;
Best Local Similarity 38.3%; Pred. No. 25;
Matches 18; Conservative 5; Mismatches 18; Indels 6; Gaps 2;

QY 4 HISSPFKYPHTQEAQKAQRSIGEMPGRLHSGMSLAICLVPLVREG 50
Db 490 HVESAWK-DVNGEFQKPT-----EMPTVLNRLNRLARVMDVLVREG 530

RESULT 12
RXRA_BRARE STANDARD; PRT; 441 AA.
ID RXRA_BRARE
```

```
AC Q90416;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE RETINOIC ACID RECEPTOR RXR-ALPHA.
GN RXRA OR NR2B1 OR RXR.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96009547; PubMed=7565671;
RA Jones B.B., Ohno C.K., Allenby G., Boffa M.B., Levin A.A.,
RA Grippo J.F., Petkovich M.;
RT "New retinoid X receptor subtypes in zebra fish (Danio rerio)
RT differentially modulate transcription and do not bind 9-cis retinoic
RT acid.";
RL Mol. Cell. Biol. 15:5226-5234(1995).
CC -!- FUNCTION: INVOLVED IN RETINOIC ACID RESPONSE PATHWAY. BINDS
CC 9-CIS RETINOIC ACID (9C-RA) (BY SIMILARITY).
CC -!- SUBUNIT: HOMODIMER OR HETERODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC NR2 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U29940; AAC59720.1; -.
DR HSSP; P19793; 2NLL.
DR ZFIN; ZDB-GENE-980526-36; rxra.
DR InterPro; IPR000003; -.
DR InterPro; IPR000536; -.
DR InterPro; IPR001628; -.
DR InterPro; IPR001723; -.
DR Pfam; PF00104; hormone_rec; 1.
DR PRINTS; PR00105; zf-C4; 1.
DR PRINTS; PR00047; STROIDFINGER.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00545; RETINOIDXR.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger; Multigene family.
FT DOMAIN 1 116 MODULATING (BY SIMILARITY).
FT DNA_BIND 117 182 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 117 137 C4-TYPE.
FT ZN_FING 153 177 C4-TYPE.
FT DOMAIN 183 206 HINGE.
FT DOMAIN 207 441 LIGAND-BINDING (BY SIMILARITY).
SQ SEQUENCE 441 AA; 48690 MW; E0F8F1EBE31CAEC CRC64;

Query Match 19.6%; Score 52.5; DB 1; Length 441;
Best Local Similarity 31.1%; Pred. No. 23;
Matches 14; Conservative 6; Mismatches 18; Indels 7; Gaps 1;

QY 2 FLHISPFKYPHTQEAQKAQRSIGEMPGRLHSGMSLS 39
Db 7 YLHLSLSLQVAHGLSLSPSPQPLGSMVSHHHHPHSLGLSPYSV 51

RESULT 13
RIR2_TRYBB STANDARD; PRT; 337 AA.
ID RIR2_TRYBB
```

AC 015910; 015880;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE SMALL CHAIN (EC 1.17.4.1)
 DE (RIBONUCLEOTIDE REDUCTASE R2 SUBUNIT).
 GN RNR2 OR NRDB.
 OS Trypanosoma brucei brucei.
 OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
 OC NCBI_TaxID=5702;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=427;
 RC MEDLINE=97338127; PubMed=9192674;
 RX Hofer A., Schmidt P.P., Graslund A., Thelander L.;
 RA "Cloning and characterization of the R1 and R2 subunits of
 RT ribonucleotide reductase from Trypanosoma brucei";
 RL Proc. Natl. Acad. Sci. U.S.A. 94:6959-6964(1997).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=97459778; PubMed=9315738;
 RX Dormeyer M., Schoneck R., Dittmar G.A.G., Krauth-Siegel R.L.;
 RA "Cloning, sequencing and expression of ribonucleotide reductase R2
 RT from Trypanosoma brucei";
 RL FEBS Lett. 414:449-453(1997).
 CC -!- FUNCTION: PROVIDES THE PRECURSORS NECESSARY FOR DNA SYNTHESIS.
 CC -!- CATALYTIC ACTIVITY: 2'DEOXYRIBONUCLEOSIDE DIPHOSPHATE + OXIDIZED
 CC THIOREDOXIN + H(2)O -> RIBONUCLEOSIDE DIPHOSPHATE + REDUCED
 CC THIOREDOXIN.
 CC -!- COFACTOR: CONTAINS TWO IRON IONS.
 CC -!- PATHWAY: FIRST REACTION IN THE DNA REPLICATION PATHWAY.
 CC -!- SUBUNIT: HETERODIMER OF A LARGE AND A SMALL CHAIN.
 CC -!- SIMILARITY: BELONGS TO THE RIBONUCLEOSIDE DIPHOSPHATE REDUCTASE
 CC SMALL CHAIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U80911; AAB70705.1; -
 DR EMBL; Y10768; CAA71741.1; -
 DR InterPro: IPR000358; -
 DR Pfam: PF00268; ribonuc_red.1.
 DR PROSITE: PS00368; RIBORED_SMALL; 1.
 KW Oxidoreductase; DNA replication; Iron.
 FT METAL 85 85 IRON 1 (BY SIMILARITY).
 FT METAL 116 116 IRON 1 AND 2 (BY SIMILARITY).
 FT METAL 119 119 IRON 1 (BY SIMILARITY).
 FT METAL 178 178 IRON 2 (BY SIMILARITY).
 FT METAL 212 212 IRON 2 (BY SIMILARITY).
 FT METAL 215 215 IRON 2 (BY SIMILARITY).
 FT ACT_SITE 123 123 BY SIMILARITY.
 FT CONFLICT 8 10 CSR -> RSA (IN REF. 2).
 FT CONFLICT 43 43 E -> K (IN REF. 2).
 SQ SEQUENCE 337 AA; 39008 MW; BD39919280C65CAE CRC64;
 Query Match 19.4%; Score 52; DB 1; Length 337;
 Best Local Similarity 31.4%; Pred. No. 20;
 Matches 11; Conservative 8; Mismatches 16; Indels 0; Gaps 0;
 QY 4 HISSPKYPTHTQEAQKRAQSRSLGMPGRHLGSSMS 38
 DB 27 YVIFPKYPIWOKYKEASSIWTVEIDLGNDMT 61
 RESULT 14
 MAPA_RAT
 ID MAPA_RAT STANDARD; PRT; 2774 AA.

AC P34926;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE MICROTUBULE-ASSOCIATED PROTEIN 1A [CONTAINS: MAP1 LIGHT CHAIN LC2].
 GN MAP1A.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP TISSUE=Brain;
 RC MEDLINE=92355629; PubMed=1379599;
 RX Langkopf A., Hammarback J.A., Mueller R., Vallee R.B., Garner C.C.;
 RA "Microtubule-associated proteins 1A and LC2. Two proteins encoded in
 RT one messenger RNA";
 RL J. Biol. Chem. 267:16561-16566(1992).
 CC -!- FUNCTION: STRUCTURAL PROTEIN INVOLVED IN THE FILAMENTOUS
 CC CROSS-BRIDGING BETWEEN MICROTUBULES AND OTHER SKELETAL ELEMENTS.
 CC -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
 CC WITH MAP1A AND MAP1B PROTEINS.
 CC -!- TISSUE SPECIFICITY: BRAIN, HEART AND MUSCLE.
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED LATE DURING NEURONAL DEVELOPMENT
 CC APPEARING WHEN AXONS AND DENDRITES BEGIN TO SOLIDIFY AND STABILIZE
 CC THEIR MORPHOLOGY.
 CC -!- DOMAIN: THE BASIC REGION CONTAINING THE REPEATS MAY BE RESPONSIBLE
 CC FOR THE BINDING OF MAP1A TO MICROTUBULES.
 CC -!- PTM: VARIOUS SERINE RESIDUES MAY BE PHOSPHORYLATED BY CAMP KINASE.
 CC -!- PTM: LC2 IS COEXPRESSED WITH MAP1A. IT IS A POLYPEPTIDE GENERATED
 CC FROM MAP1A BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
 CC BOTH MAP1A AND MAP1B.
 CC -!- SIMILARITY: TO MAP1B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M83196; AAB48069.1; -
 DR PIR; A43359; A43359.
 KW Microtubules; Repeat; Phosphorylation.
 FT CHAIN 22465 2774 MAP1 LIGHT CHAIN LC2.
 FT DOMAIN 309 496 LYS-RICH (BASIC).
 FT DOMAIN 336 541 11 X 3 AA REPEATS OF K-K-[DE].
 FT REPEAT 336 338 1.
 FT REPEAT 415 417 2.
 FT REPEAT 420 422 3.
 FT REPEAT 424 426 4.
 FT REPEAT 427 429 5.
 FT REPEAT 431 433 6.
 FT REPEAT 436 438 7.
 FT REPEAT 440 442 8.
 FT REPEAT 444 446 9.
 FT REPEAT 449 451 10.
 FT REPEAT 539 541 11.
 SQ SEQUENCE 2774 AA; 299526 MW; 3DEF74427BA9D7D7 CRC64;
 Query Match 19.4%; Score 52; DB 1; Length 2774;
 Best Local Similarity 46.2%; Pred. No. 1.8e+02;
 Matches 12; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
 QY 9 FKYPHTQEAQKRAQSRSLGMPGRHLG 34
 DB 678 FYQKHTQEAQKRAQSRSLGMPGRHLG 703
 RESULT 15
 AQPL_HUMAN

ID AC AQP1_HUMAN STANDARD; PRT; 269 AA.
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE AQUAPORIN-CHIP (WATER CHANNEL PROTEIN FOR RED BLOOD CELLS AND KIDNEY
DE PROXIMAL TUBULE) (AQUAPORIN 1) (AQP-1) (URINE WATER CHANNEL).
GN AQP1 OR CHIP28.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=92107900; PubMed=1722319;
RA Preston G.M., Agre P.;
RT "Isolation of the cDNA for erythrocyte integral membrane protein of
RT 28 kilodaltons: member of an ancient channel family.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:11110-11114(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93340184; PubMed=8340403;
RA Moon C., Preston G.M., Griffin C.A., Jabs E.W., Agre P.;
RT "The human aquaporin-CHIP gene. Structure, organization, and
RT chromosomal localization.";
RL J. Biol. Chem. 268:15772-15778(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Retinal pigment epithelium;
RX MEDLINE=96326579; PubMed=8703970;
RA Ruiz A.C., Bok D.;
RT "Characterization of the 3' UTR sequence encoded by the AQP-1 gene in
RT human retinal pigment epithelium.";
RL Biochim. Biophys. Acta 1282:174-178(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RX MEDLINE=94290349; PubMed=7517253;
RA Li X., Yu H., Koide S.S.;
RT "The water channel gene in human uterus.";
RL Biochem. Mol. Biol. Int. 32:371-377(1994).
RN [5]
RP SEQUENCE OF 1-128 FROM N.A.
RA Dempsey S., Lacy M., Holmes A., Nguyen C.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 129-269 FROM N.A.
RA Andrews S., Dubbeide C., Ryan E.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [7]
RP FUNCTION.
RX MEDLINE=92229472; PubMed=1373524;
RA Preston G.M., Carroll T.P., Guggino W.B., Agre P.;
RT "Appearance of water channels in Xenopus oocytes expressing red cell
RT CHIP28 protein.";
RL Science 256:385-387(1992).
RN [8]
RP TARGET OF MERCURY INHIBITION.
RX MEDLINE=93106996; PubMed=7677994;
RA Preston G.M., Jung J.S., Guggino W.B., Agre P.;
RT "The mercury-sensitive residue at cysteine 189 in the CHIP28 water
RT channel.";
RL J. Biol. Chem. 268:17-20(1993).
RN [9]
RP TOPOLOGY.
RX MEDLINE=94124503; PubMed=7507481;
RA Preston G.M., Jung J.S., Guggino W.B., Agre P.;
RT "Membrane topology of aquaporin CHIP. Analysis of functional epitope-
RT scanning mutants by vectorial proteolysis.";
RL J. Biol. Chem. 269:1668-1673(1994).
RN [10]
RP STRUCTURE BY ELECTRON CRYO-MICROSCOPY.
RX MEDLINE=94313979; PubMed=7518771;

RA Walz T., Smith B.L., Agre P., Engel A.;
RT "The three-dimensional structure of human erythrocyte aquaporin
RT CHIP.";
RL EMBO J. 13:2985-2993(1994).
RN [11]
RP STRUCTURE BY ELECTRON CRYO-MICROSCOPY.
RX MEDLINE=973220502; PubMed=9177353;
RA Walz T., Hirai T., Murata K., Heymann J.B., Mitsuoaka K.,
RA Fujiyoshi Y., Smith B.L., Agre P., Engel A.;
RT "The three-dimensional structure of aquaporin-1.";
RL Nature 387:624-627(1997).
RN [12]
RP X-RAY CRYSTALLOGRAPHY (3.8 ANGSTROMS).
RX MEDLINE=20487015; PubMed=11034202;
RA Murata K., Mitsuoaka K., Hirai T., Walz T., Agre P., Heymann J.B.,
RA Engel A., Fujiyoshi Y.;
RT "Structural determinants of water permeation through aquaporin-1.";
RL Nature 407:599-605(2000).
RN [13]
RP VARIANT BLOOD GROUP COLTON.
RX MEDLINE=94365170; PubMed=7521882;
RA Smith B.L., Preston G.M., Spring F., Anstee D.J., Agre P.;
RT "Human red cell aquaporin CHIP. I. Molecular characterization of ABH
RT and Colton blood group antigens.";
RL J. Clin. Invest. 94:1043-1049(1994).
RN [14]
RP FUNCTION: FORMS A WATER-SPECIFIC CHANNEL THAT PROVIDES THE PLASMA
RC MEMBRANES OF RED CELLS AND KIDNEY PROXIMAL TUBULES WITH HIGH
CC PERMEABILITY TO WATER, THEREBY PERMITTING WATER TO MOVE IN THE
CC DIRECTION OF AN OSMOTIC GRADIENT.
CC [1] SUBUNIT: HOMOTETRAMER.
CC [2] SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC [3] TISSUE SPECIFICITY: EXPRESSED IN A NUMBER OF TISSUES INCLUDING
CC ERYTHROCYTES, RENAL TUBULES, RETINAL PIGMENT EPITHELIUM,
CC HEART LUNG, SKELETAL MUSCLE, KIDNEY AND PANCREAS. WEAKLY
CC EXPRESSED IN BRAIN, PLACENTA AND LIVER.
CC [4] DOMAIN: HB AND HE TOGETHER FORM A TRANSMEMBRANE HELIX. THEY ALSO
CC FUNCTION AS PORE HELICES INVOLVED IN DEFINING SUBSTRATE
CC SPECIFICITY. THE TWO NPA REPEATS ARE SITUATED AT THE END OF HB AND
CC HE AND ARE IMPORTANT FOR THE INTERACTION BETWEEN THESE TWO
CC HELICES.
CC [5] POLYMORPHISM: AQP1 IS RESPONSIBLE FOR THE COLTON BLOOD GROUP
CC SYSTEM. APPROXIMATELY 92% OF CAUCASIANS ARE CO(A+B-) (ALA-45),
CC APPROXIMATELY 8% ARE CO(A+B+), AND ONLY 0.2% ARE CO(A-B+) (VAL-
CC 45). CO(A-B-) WHICH IS VERY RARE, IS DUE TO A COMPLETE ABSENCE OF
CC AQP1.
CC [6] MISCELLANEOUS: PHARMACOLOGICALLY INHIBITED BY SUBMILLIMOLAR
CC CONCENTRATIONS OF Hg2+.
CC [7] SIMILARITY: BELONGS TO THE TRANSMEMBRANE CHANNEL MIP FAMILY.
CC [8] This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC [9] EMBL: M77829; AAA58425.1;
CC [10] EMBL: U41517; AAC50648.1;
CC [11] EMBL: U41518; AAC50649.1;
CC [12] EMBL: S73482; AAB31193.1;
CC [13] EMBL: AC004691; AAC16481.1;
CC [14] EMBL: AC005155; AAC23788.1;
CC [15] PIR: A41616; A41616
CC [16] PDB: 1FOY; 18-OCT-00.
CC [17] MIM: 107776;
CC [18] MIM: 110450;
CC [19] InterPro: IPR000425;
CC [20] Pfam: PF00230; MIP; 1.
CC [21] PRINTS: PR00783; MINTRINSCP.
CC [22] PROSITE: PS00221; MIP; 1.
KW Transport; Transmembrane; Polymorphism; Blood group antigen;
KW Glycoprotein; 3D-structure.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 16:52:14 ; Search time 26.94 Seconds
(without alignments)
144.206 Million cell updates/sec

Title: US-09-402-713A-7

Perfect score: 268

Sequence: 1 MFLHSSPFKYPHTQAEK.....HLGSSMSLALCLVPLVREGH 51

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	62.5	23.3	441	2 E86758	dihydroorotase (EC
2	60.5	22.6	1239	2 T13809	probable disintegr
3	60	22.4	187	2 B81903	hypothetical prote
4	60	22.4	218	2 C81118	hypothetical prote
5	56.5	21.1	1061	2 C75508	acriflavin resista
6	56	20.9	338	2 C75459	probable endonucle
7	55	20.5	1039	2 A41131	lymphocyte-Peyer's
8	54.5	20.3	244	2 A83236	hypothetical prote
9	54.5	20.3	305	2 T47844	hypothetical prote
10	54.5	20.3	461	2 S65665	actin-binding prot
11	54.5	20.3	461	2 S65666	actin-binding prot
12	54	20.1	204	2 T51012	hypothetical prote
13	54	20.1	257	2 G72081	conserved hypothet
14	54	20.1	257	2 F86540	Cn254 hypothetical
15	54	20.1	633	2 B82990	hypothetical prote
16	53.5	20.0	348	2 T14128	NADH dehydrogenase
17	53.5	20.0	446	2 JC4989	glucuronate permease
18	53.5	20.0	446	2 E86010	low affinity gluco
19	53.5	20.0	449	2 D72127	hypothetical prote
20	53.5	20.0	451	2 C49898	cellobiose phospho
21	52.5	19.6	216	2 E75446	amino acid ABC tra
22	52.5	19.6	357	2 E83632	hypothetical prote
23	52.5	19.6	441	2 T50515	retinoid X recepto
24	52.5	19.6	493	2 T43465	hypothetical prote
25	52.5	19.6	563	2 T17316	hypothetical prote
26	52	19.4	210	2 H75599	probable acetyltra
27	52	19.4	493	2 F96696	protein FIN21.12 {
28	52	19.4	2774	2 A43359	microtubule-associ
29	51.5	19.2	173	2 D83486	cobinamide kinase

30 51.5 19.2 200 2 H84715 probable phytoeyan
31 51.5 19.2 269 2 A41616 erythrocyte integr
32 51.5 19.2 380 2 S75929 sensory transducti
33 51.5 19.2 449 2 C86496 hypothetical prote
34 51.5 19.2 449 2 A81544 hypothetical prote
35 51.5 19.2 470 1 RKMWLX ribulose-bisphosph
36 51.5 19.2 569 2 T03390 4-comumarate--CoA 1
37 51 19.0 137 2 B72786 hypothetical prote
38 51 19.0 464 2 T45612 N-hydroxycinnamoyl
39 51 19.0 633 2 A75069 dipeptide transpor
40 51 19.0 1038 2 S06046 integrin alpha-4 c
41 51 19.0 2137 2 T05244 hypothetical prote
42 51 19.0 2492 1 A44213 nonstructural poly
43 51 19.0 2492 1 C44213 nonstructural poly
44 50.5 18.8 115 2 G72617 hypothetical prote
45 50.5 18.8 409 1 A47527 transcription fact

ALIGNMENTS

RESULT 1

E86758

dihydroorotase (EC 3.5.2.3) [imported] - Lactococcus lactis subsp. lactis (strain ILI
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 31-Mar-2001
C:Accession: E86758

R:Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Eh

Genome Res. in press, 2001

A:Title: The complete genome sequence of the lactic acid bacterium.

A:Reference number: A86625

A:Accession: E86758

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-441 <STO>

A:Cross-references: GB:AB005176; NID:g12724023; PIDN:AAK05167.1; GSPDB:GN00146

A:Experimental source: strain ILI403

C:Genetics:

A:Gene: pyrC

C:Superfamily: Bacillus dihydroorotase; Bacillus dihydroorotase homology

C:Keywords: hydrolase

Query Match

Best Local Similarity 23.3%; Score 62.5; DB 2; Length 441;

Matches 17; Conservative 4; Mismatches 16; Indels 3; Gaps 1;

QY 12 PHTQAEQAEQSRSLGEMFGRHLGSSMSLALCLVPLVREGH 51

DB 325 PHTRE---EKNVSLDKAPSGMIGLETSLQLGLTLNVAKGH 361

RESULT 2

T13809

probable disintegrin (EC 3.4.24.-) - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000

C:Accession: T13809

R:Rooken, J.; Pan, D.; Xu, T.; Rubin, G.M.

Science 273, 1227-1231, 1996

A:Title: KUZ, a conserved metalloprotease-disintegrin with two roles during Drosophil

A:Reference number: 217772; MUID:96355647

A:Accession: T13809

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1239 <ROO>

A:Cross-references: EMBL:U60591; NID:g1531632; PID:g1531633; PIDN:AAK47275.1

C:Genetics:

A:Gene: kuz

A:Cross-references: FlyBase:FBgn0015954

A:Map position: 2

C:Keywords: hydrolase; metalloproteinase

DD 89 LFLSLRGEIVFPKNETAESEFEKSVHER--RQEGNAGSGRKKQLLDLVRRGH 137

RESULT 5

acriflavin resistance protein acrF - *Synechocystis* sp. (strain PCC 6803)
N:Alternate names: protein sir2131
C:Species: *Synechocystis* sp.

C; Date: 23-Apr-1997 #sequence_revision 23-Apr-1997 #text_change 20-Jun-2000
C; Accession: S75508
R; Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
O. K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A; Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocys*
s.

Query Match	21.1%	Score 56.5;	DB 2;	Length 1061;
Best Local Similarity	31.0%;	Pred. No. 28;		
Matches 13:	Conservative	11;	Mismatches 17;	Indels 11;
				Gaps 1;

RESULT 6
C75459
probable endonuclease III - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: C75459

Query Match	20.9%	Score 56;	DB 2;	Length 338;
Best Local Similarity	38.1%;	Pred. No. 9.3;		
Matches	8;	Conservative	8;	Mismatches
			5;	Indels
			0;	Gaps

RESULT 7
A41131
lymphocyte-Peyer's patch adhesion molecule alpha 4 precursor - mouse
N;Alternate names: integrin alpha-4

A;Note: T209.l80
C;Superfamily: Arabidopsis thaliana hypothetical protein At2g44600

Query Match 20.3%; Score 54.5; DB 2; Length 305;
Best Local Similarity 42.5%; Pred. No. 13;
Matches 17; Conservative 7; Mismatches 9; Indels 7; Gaps 3;

QY 16 EAOKEAQRS-LGEMPGRRH-----IGSSMS-LALCLVPLVR 48
| : | : : ||| : ||| : ||| ||||
Db 210 ESIELRRTPATKTPGRKRKIAMGIGKSMGMATCLSPLVR 249

RESULT 10
S65665
actin-binding protein p57 - human
C;Species: Homo sapiens (man)
C;Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 21-Jul-2000
C;Accession: S65665
R;Suzuki, K.; Nishihata, J.; Arai, Y.; Honma, N.; Yamamoto, K.; Irimura, T.; Toyoshima FBS Lett. 364, 283-288, 1995
A;Title: Molecular cloning of a novel actin-binding protein, p57, with a WD repeat an A;Reference number: S65614; MUID:95278344
A;Accession: S65665
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-461 <SUZ>
A;Cross-references: EMBL:D44497; NID:g927648; PIDN:BAA07940.1; PID:g927649
C;Genetics:
A;Gene: GDB:p57
A;Cross-references: GDB:686252
C;Superfamily: unassigned WD repeat proteins; WD repeat homology
F;77-111/Domain: WD repeat homology <WDR>

Query Match 20.3%; Score 54.5; DB 2; Length 461;
Best Local Similarity 43.8%; Pred. No. 21;
Matches 14; Conservative 3; Mismatches 8; Indels 7; Gaps 1;

QY 2 FLHTSPFKYPHTQEAOKEAQRSLGEMPGRHLL 33
| | | | | : | | | | : | | | | |
Db 303 FLHYLSMF-----SSKSQSGMGYMPKRGRL 327

RESULT 11
S65666
actin-binding protein p57 - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 26-May-2000
C;Accession: S65666; S65614
R;Suzuki, K.; Nishihata, J.; Arai, Y.; Honma, N.; Yamamoto, K.; Irimura, T.; Toyoshima FBS Lett. 364, 283-288, 1995
A;Title: Molecular cloning of a novel actin-binding protein, p57, with a WD repeat an A;Reference number: S65614; MUID:95278344
A;Accession: S65666
A;Molecule type: mRNA
A;Residues: 1-461 <SUZ>
A;Cross-references: EMBL:D44496; NID:g927646; PIDN:BAA07939.1; PID:d1008527; PID:g927649
A;Experimental source: spleen
A;Accession: S65614
A;Molecule type: protein
A;Residues: 21-43;215-230;234-245;334-339;416-430 <SU2>
C;Superfamily: unassigned WD repeat proteins; WD repeat homology
C;Keywords: actin binding; leucine zipper
F;77-111/Domain: WD repeat homology <WD1>
F;127-161/Domain: WD repeat homology <WD2>
F;433-454/Region: leucine zipper motif

Query Match 20.3%; Score 54.5; DB 2; Length 461;
Best Local Similarity 43.8%; Pred. No. 21;
Matches 14; Conservative 3; Mismatches 8; Indels 7; Gaps 1;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 16:50:50 ; Search time 26.85 seconds
(without alignments)
39.110 Million cell updates/sec

Title: US-09-402-713A-7
Perfect score: 268
Sequence: 1 MFLHISPPKYPHTQEAQKE.....HLGSSMSLALCLVPLVREGH 51

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA: *
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	60.5	22.6	1239	2	US-08-937-931-2
2	60.5	22.6	1239	4	US-09-285-502-2
3	53	19.8	1151	1	US-08-286-889-37
4	53	19.8	1151	1	US-08-485-618-37
5	53	19.8	1151	1	US-08-362-652-37
6	53	19.8	1151	2	US-08-605-672-37
7	53	19.8	1151	2	US-08-482-293A-37
8	53	19.8	1151	2	US-08-943-363-37
9	53	19.8	1151	4	US-09-193-043-37
10	53	19.8	1161	1	US-08-485-618-55
11	53	19.8	1161	1	US-08-362-652-55
12	53	19.8	1161	2	US-08-605-672-55
13	53	19.8	1161	2	US-08-482-293A-55
14	53	19.8	1161	2	US-08-943-363-55
15	53	19.8	1161	4	US-09-193-043-55
16	52.5	19.6	685	3	US-09-031-563-21
17	52.5	19.6	1055	3	US-09-031-563-27
18	52.5	19.6	1315	3	US-09-031-563-2
19	52.5	19.6	1315	3	US-09-031-563-25
20	51.5	19.2	269	1	US-08-447-554-5
21	51.5	19.2	269	1	US-08-468-763-17
22	51.5	19.2	269	1	US-08-448-160-5
23	51.5	19.2	269	2	US-08-393-996A-17
24	51	19.0	1155	1	US-08-286-889-46
25	51	19.0	1155	1	US-08-485-618-46
26	51	19.0	1155	1	US-08-362-652-46
27	51	19.0	1155	2	US-08-605-672-46

28	51	19.0	1155	2	US-08-482-293A-46	Sequence 46, Appl
29	51	19.0	1155	2	US-08-943-363-46	Sequence 46, Appl
30	51	19.0	1155	4	US-09-193-043-46	Sequence 46, Appl
31	51	19.0	1161	1	US-08-173-497-2	Sequence 2, Appli
32	51	19.0	1161	1	US-08-286-889-2	Sequence 2, Appli
33	51	19.0	1161	1	US-08-485-618-2	Sequence 2, Appli
34	51	19.0	1161	1	US-08-485-618-53	Sequence 53, Appl
35	51	19.0	1161	1	US-08-485-618-99	Sequence 99, Appl
36	51	19.0	1161	1	US-08-362-652-2	Sequence 2, Appli
37	51	19.0	1161	1	US-08-362-652-53	Sequence 53, Appl
38	51	19.0	1161	2	US-08-605-672-2	Sequence 2, Appli
39	51	19.0	1161	2	US-08-605-672-53	Sequence 53, Appl
40	51	19.0	1161	2	US-08-605-672-99	Sequence 99, Appl
41	51	19.0	1161	2	US-08-482-293A-2	Sequence 2, Appli
42	51	19.0	1161	2	US-08-482-293A-53	Sequence 53, Appl
43	51	19.0	1161	2	US-08-482-293A-99	Sequence 99, Appl
44	51	19.0	1161	2	US-08-943-363-2	Sequence 2, Appli
45	51	19.0	1161	2	US-08-943-363-99	Sequence 99, Appl

ALIGNMENTS

RESULT 1
US-08-937-931-2
; Sequence 2, Application US/08937931
; Patent No. 5935792
; GENERAL INFORMATION:
; APPLICANT: Rudin, Gerald M.
; APPLICANT: Pan, Duojia
; APPLICANT: Rooke, Jenny
; APPLICANT: Yavari, Reza
; APPLICANT: Xu, Tian
; TITLE OF INVENTION: KUZ: A No. 5935792el Family of Metalloproteases
; NUMBER OF INVENTIONS: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,931
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B97-081
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1239 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-937-931-2

Query Match 22.6%; Score 60.5; DB 2; Length 1239;
Best Local Similarity 47.2%; Pred. No. 2.9;
Matches 17; Conservative 4; Mismatches 10; Indels 5; Gaps 2;
QY 9 FKYPHTQEAQKQRSIGEM----PGRHLGSSMSLA 40


```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32797
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-485-618-37

Query Match 19.8%; Score 53; DB 1; Length 1151;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

QY 14 TQEAQKEAQRSLGEMPGRHLSGSSMSIALCLVPLVREG 50
      ||||: :| :| :|| :|| :|| :|| :||
Db 428 TQEARHWRPKS--EVRGTQIGSYFGASLCSVDVDRDG 462

RESULT 5
US-08-362-652-37
; Sequence 37, Application US/08362652
; Patent No. 5766850
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,652
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32391
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
```

```
; MOLECULE TYPE: protein
; US-08-362-652-37

Query Match 19.8%; Score 53; DB 1; Length 1151;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

QY 14 TQEAQKEAQRSLGEMPGRHLSGSSMSIALCLVPLVREG 50
      ||||: :| :| :|| :|| :|| :|| :||
Db 428 TQEARHWRPKS--EVRGTQIGSYFGASLCSVDVDRDG 462

RESULT 6
US-08-605-672-37
; Sequence 37, Application US/08605672
; Patent No. 5817515
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/605,672
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32684
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-605-672-37

Query Match 19.8%; Score 53; DB 2; Length 1151;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

QY 14 TQEAQKEAQRSLGEMPGRHLSGSSMSIALCLVPLVREG 50
      ||||: :| :| :|| :|| :|| :|| :||
Db 428 TQEARHWRPKS--EVRGTQIGSYFGASLCSVDVDRDG 462
```

```
RESULT 7
US-08-482-293A-37
; Sequence 37, Application US/08482293A
; Patent No. 5831029
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5831029el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/482,293A
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32684
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-482-293A-37

Query Match 19.8%; Score 53; DB 2; Length 1151;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

QY 14 TQBAQKQKSLGEMPGRLHSGSMALCLVPLVREG 50
||||: | | | | | | | | | | | | | | | | |
Db 428 TQEAHRWRPKS--EVGRGTQIGSYFGASLCSDVDVDRDG 462

RESULT 8
US-08-943-363-37
; Sequence 37, Application US/08943363
; Patent No. 5837478
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5837478el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
```

```
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943,363
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32684
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-943-363-37

Query Match 19.8%; Score 53; DB 2; Length 1151;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

QY 14 TQBAQKQKSLGEMPGRLHSGSMALCLVPLVREG 50
||||: | | | | | | | | | | | | | | | | |
Db 428 TQEAHRWRPKS--EVGRGTQIGSYFGASLCSDVDVDRDG 462

RESULT 9
US-09-193-043-37
; Sequence 37, Application US/09193043
; Patent No. 6251395
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 6251395el Human 2
; FILE REFERENCE: 27866/35004
; CURRENT APPLICATION NUMBER: US/09/193,043
; CURRENT FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: 08/173,497
; EARLIER FILING DATE: 1993-12-23
; EARLIER APPLICATION NUMBER: 08/286,889
; EARLIER FILING DATE: 1994-08-05
; EARLIER APPLICATION NUMBER: 08/362,652
; EARLIER FILING DATE: 1994-12-21
; EARLIER APPLICATION NUMBER: 08/943,363
; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
; LENGTH: 1151
; TYPE: PRT
; ORGANISM: Rattus rattus
US-09-193-043-37
```

RESULT 11
US-08-362-652-55

```

US-08-605-672-55
; Sequence 55, Application US/08605672
; Patent No. 5817515
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5817515e
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESS: Marshall, O'Toole,
; STREET: 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

Query Match 19.8%; Score 53; DB 1; Length 1161;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14: Conservative 7; Mismatches 14: Indels 2: Gaps 1:

Query Match 19.8%; Score 53; DB 2; Length 1161;
Best Local Similarity 37.8%; Pred. NO. 34;
Matches 14; Conservative 7; Mismatches 14; Indels

Qy 14 TQEAEQAQRSLGEMPRHILGSSMSLALCLVPLVREG 50
||||| : | : ||| : ||| : | : |
Dd 438 TQEARHWKPKS--EVRGTIGSYFGASLCSDVDVRDG 472

RESULT 14

US-08-943-303-53
; Sequence 55, Application US/08943363
; Patent No. 5837478

; TITLE OF INVENTION: No. 5837478el Human 2 Integrin Alpha Subunit
 ;
 ; NUMBER OF SEQUENCES: 114
 ;
 ; CORRESPONDENCE ADDRESS:
 ;
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 ;
 ; STREET: 233 South Wacker Drive, 6300 Sear Tower
 ;
 ; CITY: Chicago

RESULT 13
US-08-482-293A-55

```

US 06 462 293A 33
; Sequence 55, Application US/0842293A
; Patent No. 5831029
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Gallatin, W. Michael
; TITLE OF INVENTION: No. 5831029el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103

```

```

:
:
: GENERAL INFORMATION:
:
: * APPLICANT: Gallatin, W. Michael
:
: * APPLICANT: Van der Vieren, Monica
:
: TITLE OF INVENTION: No. 5831029e1 Human 2 Integrin Alpha
:
: NUMBER OF SEQUENCES: 103
:
: CORRESPONDENCE ADDRESS:
:
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
:
: STREET: 233 South Wacker Drive, 6300 Sear Tower
:
:

```

CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402

```

:
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/482,293A
:
:

```

```

; FILING DATE:
; CLASSIFICATION: 530
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
;

```

APPLICATION NUMBER: US 08/173,497
 FILING DATE: 23-DEC-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/286,889
 FILING DATE: 5-AUG-1994

; APPLICATION NUMBER: US 08/286,889
 ; FILING DATE: 5-AUG-1994

; LENGTH: 1161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-943-363-55

Query Match 19.8%; Score 53; DB 2; Length 1161;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

Qy 14 TOEAQKEAQRSLGEMPGRHLSGSSMSLALCLVPLVREG 50
||||: :| |:| :||| :||| :|:
Db 438 TQEARHWRPKS--EVRGTQIGSYFGASLCSVDVDRDG 472

RESULT 15

US-09-193-043-55
; Sequence 55, Application US/09193043
; Patent No. 6251395

; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.

; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: NO. 6251395el Human 2

; FILE REFERENCE: 27866/35004
; CURRENT APPLICATION NUMBER: US/09/193,043

; CURRENT FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: 08/173,497

; EARLIER FILING DATE: 1993-12-23
; EARLIER APPLICATION NUMBER: 08/286,889

; EARLIER FILING DATE: 1994-08-05
; EARLIER APPLICATION NUMBER: 08/362,652

; EARLIER FILING DATE: 1994-12-21
; EARLIER APPLICATION NUMBER: 08/943,363

; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 114

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 55

; LENGTH: 1161
; TYPE: PRT

; ORGANISM: Rattus rattus
US-09-193-043-55

Query Match 19.8%; Score 53; DB 4; Length 1161;
Best Local Similarity 37.8%; Pred. No. 34;
Matches 14; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

Qy 14 TOEAQKEAQRSLGEMPGRHLSGSSMSLALCLVPLVREG 50
||||: :| |:| :||| :||| :|:
Db 438 TQEARHWRPKS--EVRGTQIGSYFGASLCSVDVDRDG 472

Search completed: July 30, 2001, 16:50:51
Job time: 166 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 16:51:40 ; Search time 42.83 seconds
(without alignments)
72.188 Million cell updates/sec

Title: US-09-402-713a-7
Perfect score: 268
Sequence: 1 MFLHTSSPPKYPHTQBAQKE.....HLGSSMSLALCLVPLVREGH 51

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 50623988 residues
Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0601.*
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	268	100.0	51	AAW79736	Prostate cancer an
2	268	100.0	51	AAW79738	Prostate cancer an
3	77	28.7	16	AAV07887	Human secreted pro
4	63	23.5	514	AAV87343	Human signal pepti
5	63	23.5	602	AAV43919	Human cancer assoc
6	60.5	22.6	1239	AAV56131	Drosophila melanoga
7	59	22.0	897	AAV07092	Colon cancer assoc
8	59	22.0	1141	AAV43008	Human ORF3 ORP2772
9	56	20.9	11	AAW79737	Prostate cancer an
10	55	20.5	184	AAV40271	Human ORF3 ORF35 p
11	55	20.5	2432	AAV85565	Human homologue of

12	54.5	20.3	99	20	AAV29863	Human secreted pro
13	54.5	20.3	100	20	AAV25713	Human secreted pro
14	54.5	20.3	461	17	AAV98341	Human p57 protein.
15	54.5	20.3	461	17	AAV98342	Bovine p57 protein
16	54.5	20.3	598	17	AAV98344	GST/truncated huma
17	54.5	20.3	673	17	AAV98343	GST/human p57 fusi
18	54	20.1	414	21	AAV22216	Arabidopsis thalia
19	54	20.1	414	21	AAV22216	Arabidopsis thalia
20	54	20.1	450	21	AAV22215	Arabidopsis thalia
21	54	20.1	450	21	AAV22215	Arabidopsis thalia
22	54	20.1	453	21	AAV22214	Arabidopsis thalia
23	54	20.1	453	21	AAV22214	Arabidopsis thalia
24	53.5	20.0	112	20	AAV60194	Human endometrium
25	53	19.8	1151	16	AAV78179	Rat alpha-d compos
26	53	19.8	1151	18	AAV23059	Rat beta 2 integri
27	53	19.8	1151	19	AAV72834	Rat alpha-d #2. R
28	53	19.8	1151	19	AAV65101	Rat beta-integrin
29	53	19.8	1151	19	AAV60001	Rat alpha d polype
30	53	19.8	1151	20	AAV73344	Rat alpha-d protein
31	53	19.8	1151	21	AAV07371	Rat alpha-d partia
32	53	19.8	1161	16	AAV78169	Rat alpha-d subuni
33	53	19.8	1161	18	AAV23062	Rat beta 2 integri
34	53	19.8	1161	19	AAV72824	Rat alpha-d #1. R
35	53	19.8	1161	19	AAV65104	Rat beta-integrin
36	53	19.8	1161	19	AAV60004	Rat alpha d polype
37	53	19.8	1161	20	AAV73345	Rat alpha-d protein
38	53	19.8	1161	21	AAV07374	Rat alpha-d protei
39	52.5	19.6	109	20	AAV48422	Human prostate can
40	52.5	19.6	1055	20	AAV31730	Human fused protei
41	52.5	19.6	1315	20	AAV43265	Human Fused Relate
42	52.5	19.6	1315	20	AAV31725	Human fused protei
43	52.5	19.6	1315	20	AAV31729	Human fused protei
44	52.5	19.6	1315	21	AAV30679	Amino acid sequenc
45	52.5	19.6	1333	20	AAV31727	Human fused cDNA i

ALIGNMENTS

RESULT 1

AAW79736
ID AAW79736 standard; Protein; 51 AA.

XX AAW79736;

XX AC AAW79736;

XX DT 30-DEC-1998 (first entry)

XX DE Prostate cancer antigen (PCA3) protein variant 1.

XX KW Prostate cancer antigen protein variant 1; PCA3; prostatic cancer;

XX OS Homo sapiens.

XX PN WO9845420-A1.

XX PD 15-OCT-1998.

XX PF 09-APR-1998; 98WO-CA00346.

XX PR 10-APR-1997; 97US-0041836.

XX PA (DIAG-) DIAGNOCURE INC.

XX PI Bussemakers MJG;

XX XX WPI; 1998-568347/48.

XX DR N-PSDB; AAV62427.

XX PT New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,

XX PT prevention and treatment of prostatic cancer

XX PS Claim 16; Fig 2B-2J; 111pp; English.

XX The present sequence represents the prostate cancer antigen (PCA3)
 CC protein variant 1 encoded by a PCA3 spliced cDNA sequence comprising
 CC of exons 1, 2, 3, 4a and 4b of the PCA3 gene. The invention claims
 CC for PCA3 cDNA variants and the proteins they encode. The invention
 CC also claims for antibodies against PCA3 protein. The antibodies are
 CC claimed to be useful for detecting PCA3 protein in immunoassay tests,
 CC for diagnosing, assessing and prognosing of prostatic cancer (PC).
 CC Antibodies, optionally coupled to a cytotoxin or radioisotope, and
 CC nucleic acids antisense to PCA3 cDNA are claimed to be useful for
 CC treating PC, while determining elevated levels of PCA3 (as RNA or
 CC protein) is useful for detecting a predisposition to development of
 CC PC, e.g. in prenatal tests. Detecting PCA3 protein allows
 CC differentiation between malignant and benign prostatic disease,
 CC and the level of PCA3 expression allows correlation with the grade of
 CC tumour. PCA3 protein and its fragments are also claimed to be useful
 CC in vaccines for preventing PC; in drug screens for identifying
 CC specific (antagonists (potentially useful therapeutically) and for
 CC studying protein-DNA interactions.
 XX
 SQ Sequence 51 AA;

Query Match 100.0%; Score 268; DB 19; Length 51;
 Best Local Similarity 100.0%; Pred. No. 4.7e-32;
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFLHSSPFKYPHTQEAQKRAORSICGMPGRHLGSSMSLALCLVPLVREGH 51
 |||||
 Db 1 mflhsspfkyphtqeaqkrsigmpgrhlgsmslalcplvrvregh 51

RESULT 2

AAW79738
 ID AAW79738 standard; Protein; 51 AA.

XX AC AAW79738;

XX DT 30-DEC-1998 (first entry)

XX DE Prostate cancer antigen (PCA3) wild-type protein.

XX Prostate cancer antigen protein variant 1; PCA3; prostatic cancer;
 KW PC.

XX OS Homo sapiens.

XX PN W09845420-A1.

XX PD 15-OCT-1998.

XX PF 09-APR-1998; 98WO-CA00346.

XX PR 10-APR-1997; 97US-0041836.

XX PA (DIAG-) DIAGNOCURE INC.

XX PI Bussemakers MJG;

XX DR WPI; 1998-568347/48.

XX DR N-PSDB; AAV62430.

XX PT New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
 PT prevention and treatment of prostatic cancer

XX PS Claim 16; Fig 5B-5F; 111pp; English.

XX The present sequence represents the prostate cancer antigen (PCA3)
 CC wild-type protein sequence encoded by a PCA3 wild-type cDNA sequence
 CC comprising of exons 1, 2, 3, 4a-4d of the PCA3 gene. The invention
 CC claims for PCA3 cDNA variants and the proteins they encode. The
 CC invention also claims for antibodies against PCA3 protein. The
 CC antibodies are claimed to be useful for detecting PCA3 protein in

CC immunoassay tests, for diagnosing, assessing and prognosing of
 CC prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin
 CC or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed
 CC to be useful for treating PC, while determining elevated levels of
 CC PCA3 (as RNA or protein) is useful for detecting a predisposition to
 CC development of PC, e.g. in prenatal tests. Detecting PCA3 protein
 CC allows differentiation between malignant and benign prostatic disease,
 CC and the level of PCA3 expression allows correlation with the grade of
 CC tumour. PCA3 protein and its fragments are also claimed to be useful
 CC in vaccines for preventing PC; in drug screens for identifying
 CC specific (antagonists (potentially useful therapeutically) and for
 CC studying protein-DNA interactions.
 XX

SQ Sequence 51 AA;

Query Match 100.0%; Score 268; DB 19; Length 51;
 Best Local Similarity 100.0%; Pred. No. 4.7e-32;
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFLHSSPFKYPHTQEAQKRAORSICGMPGRHLGSSMSLALCLVPLVREGH 51
 |||||
 Db 1 mflhsspfkyphtqeaqkrsigmpgrhlgsmslalcplvrvregh 51

RESULT 3

AAV07887

ID AAV07887 standard; Protein; 16 AA.

XX AC AAV07887;

XX DT 06-JUL-1999 (first entry)

XX DE Human secreted protein fragment encoded from gene 36.

XX KW Human; secreted protein; treatment; prevention; protein therapy; AIDS;
 KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
 KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
 KW immune system disease; autoimmune disease; hepatic disease; lymphoma;
 KW renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
 KW cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
 KW pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
 KW arthritis; malignancy; digestive; endocrine; infection.

XX OS Homo sapiens.

XX PN W09918208-A1.

XX PD 15-APR-1999.

XX PF 01-OCT-1998; 98WO-US20775.

XX PR 02-OCT-1997; 97US-0060884.

XX PR 02-OCT-1997; 97US-0060833.

XX PR 02-OCT-1997; 97US-0060836.

XX PR 02-OCT-1997; 97US-0060837.

XX PR 02-OCT-1997; 97US-0060838.

XX PR 02-OCT-1997; 97US-0060839.

XX PR 02-OCT-1997; 97US-0060843.

XX PR 02-OCT-1997; 97US-0060862.

XX PR 02-OCT-1997; 97US-0060866.

XX PR 02-OCT-1997; 97US-0060874.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Carter KC, Duan DR, Endress GA, Feng P, Ferrie AM;

XX PI Florence KA, Greene JM, Janat F, Lafleur DW, Ni J;

XX PI Rosen CA, Ruben SM, Shi Y, Young P, Yu G;

XX DR WPI; 1999-264022/22.

XX DR N-PSDB; AAX37486.

XX PT New isolated human genes and the secreted polypeptides they encode

XX PS Claim 1b; Page 297; 368pp; English.

XX CC This invention describes novel isolated human genes and the secreted

CC proteins they encode. The products of the invention are useful for

CC preventing, treating or ameliorating medical conditions, e.g. by protein

CC or gene therapy. Also pathological conditions can be diagnosed by

CC determining the amount of the new polypeptides in a sample or by

CC determining the presence of mutations in the new polynucleotides.

CC Specific uses are described for each of the 101 polynucleotides, based

CC on which tissues they are most highly expressed in, and include

CC developing products for the diagnosis or treatment of cancer, tumours,

CC neurodegenerative disorders, developmental abnormalities and fetal

CC deficiencies, blood disorders, leukemias, diseases of the immune system,

CC autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,

CC allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate

CC disease, skeletal or cardiac muscle disorders, pulmonary disorders,

CC transplant rejection, disorders involving osteoclasts such as

CC osteoporosis, arthritis or malignancies, digestive/endocrine disorders,

CC infections and AIDS. The human secreted proteins of the invention are

CC represented in AAY07853-Y07993 and the encoding nucleic acids are

CC represented in AAX37451-X37552.

XX SQ Sequence 16 AA;

Query Match 28.7%; Score 77; DB 20; Length 16;

Best Local Similarity 100.0%; Pred. No. 0.00012; Indels 0; Gaps 0;

Matches 15; Conservative 0; Mismatches 0;

QY 37 MSLALCLVPLVREGH 51

Db 1 mslalclvplvreg 15

RESULT 4

RAY87343

ID AAY87343 standard; Protein; 514 AA.

XX AC AAY87343;

XX DT 11-MAY-2000 (first entry)

XX DE Human signal peptide containing protein HSPP-120 SEQ ID NO:120.

XX KW Human; signal peptide-containing protein; HSPP; diagnosis; cancer;

KW inflammation; cardiovascular disease; anticancer; anti-inflammatory;

KW antimicrobial; neurotropic; neuroprotective; cardiovascular; hepatotropic;

KW antiasthmatic; gene therapy; cell proliferation; neurological disorder;

KW reproductive disorder; developmental disorder; arteriosclerosis;

KW cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;

KW asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;

KW Parkinson's disease; Huntington's diseases; ovulatory defect;

XX KW muscular dystrophy.

XX OS Homo sapiens.

XX PN WO200000610-A2.

XX PD 06-JAN-2000.

XX PF 25-JUN-1999; 99WO-US14484.

XX PR 26-JUN-1998; 98US-0090762.

XX PR 31-JUL-1998; 98US-0094983.

XX PR 01-OCT-1998; 98US-0102686.

XX PR 11-DEC-1998; 98US-0112129.

XX PA (INCY-) INCYTE PHARM INC.

XX PI Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KJ, Baughn MR;

PI Akerblom IE, Au-Young J, Yue H, Patterson C, Reddy R, Hillman JL;

PI Bandman O;

XX WPI; 2000-160673/14.

DR N-PSDB; AAZ98228.

XX PT New human signal peptide-containing proteins useful in treatment,

PT prevention and diagnosis of e.g. cancer, inflammation and

PT cardiovascular disease

XX PS Claim 1; Page 239-240; 327pp; English.

XX CC AAZ98109 to AAZ98242 encode AAY87224 to AAY87357 which represent the

CC human signal peptide-containing proteins HSPP-1 to HSPP-134. HSPPs have

CC anticancer, anti-inflammatory, antimicrobial, neurotropic, hepatotropic,

CC neuroprotective, cardiovascular and antiasthmatic activities, and can

CC be used in gene therapy. HSPPs can be used to treat or prevent disorders

CC associated with decreased activity or function of HSPP. Antagonists of

CC HSPP are used to treat or prevent disorders associated with increased

CC activity or function of HSPP. Such diseases include cell proliferation

CC (including cancer), inflammation, cardiovascular, neurological,

CC reproductive or developmental disorders, (e.g. arteriosclerosis,

CC cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,

CC asthma, Crohn's disease, microbial or other infections, congestive or

CC ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's

CC diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSPP

CC nucleic acids can be used for the recombinant production of HSPP, for

CC detecting HSPP in standard hybridisation and amplification assays (for

CC diagnosis and monitoring), in gene therapy, as antisense,

CC triplex-forming or ribozyme therapeutics, for detecting related sequences

CC or genetic variations, and for chromosomal mapping. HSPP are also used to

CC raise specific antibodies (Ab) and to screen for agonists and

CC antagonists (potential therapeutic agents). Ab are used to diagnose, or

CC monitor, HSPP-related diseases (in usual immunoassays), as therapeutic

CC antagonists, in competitive drug screens, and for purification of HSPP

CC from natural sources.

XX SQ Sequence 514 AA;

Query Match 23.5%; Score 63; DB 21; Length 514;

Best Local Similarity 42.1%; Pred. No. 0.92; Indels 2; Gaps 2;

Matches 16; Conservative 7; Mismatches 7;

QY 12 PHTQEAQKEAQRSL-GEMPGRRHLGSSMSLAL-CLVPLV 47

Db 471 praedlaedsgsslygrapgrhtwslilaalacvplll 508

RESULT 5

AAB43919

ID AAB43919 standard; Protein; 602 AA.

XX AC AAB43919;

XX DT 08-FEB-2001 (first entry)

XX DE Human cancer associated protein sequence SEQ ID NO:1364.

XX KW Human; cancer associated gene; cancer antigen; detection; cancer;

KW diagnosis; cytostatic; proliferative; vulnery; immunomodulator;

KW antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;

KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;

KW dermatological; neuroprotective; thrombolytic; coagulant; neurotropic;

KW vasotopic; antipsoriatic; antiangiogenic; gene therapy; inflammation;

KW immune disorder; haematopoietic cell disorder; autoimmune disorder;

KW allergic reaction; graft versus host disease; organ rejection;

KW haemostatic; thrombolytic; cardiovascular disorder; infection;

KW neurological disease; drug screening.

XX OS Homo sapiens.

XX PN WO2000055350-A1.

XX PD 21-SEP-2000.

```
XX 08-MAR-2000; 2000WO-US05882.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI: 2000-587533/55.
XX N-PSDB; AAC78128.
XX
XX Novel isolated nucleic acids comprising sequences encoding peptides
XX useful for treating or diagnosing e.g. cancer -
XX
XX Claim 11; Page 2024-2026; 2352pp; English.
XX
XX AAC77607 to AAC78448 encode the human cancer associated proteins given
XX in AAB43398 to AAB44239. The proteins can have activities based on the
XX tissues and cells the genes are expressed in. Example of activities
XX include: cytostatic; proliferative; vulnerable; immunomodulator;
XX antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
XX antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
XX dermatological; neuroprotective; cardiac; thrombolytic; coagulant;
XX neurotropic; vasotropic; antipsoriatic and antiangiogenic. The
XX polynucleotides and polypeptides can be used for preventing, treating or
XX ameliorating medical conditions and diagnosing pathological conditions.
XX Polynucleotides, polypeptides, antibodies, agonists and antagonists from
XX the present invention may be used to treat immune disorders by activating
XX or inhibiting the proliferation, differentiation or mobilisation of
XX immune cells, to treat disorders of haematopoietic cells, autoimmune
XX disorders, allergic reactions, graft versus host disease and organ
XX rejection, modulate haemostatic or thrombolytic activity, modulate
XX inflammation, cancers, cardiovascular disorders, neurological disease and
XX bacterial or viral infections. The peptides, nucleotides, antibodies,
XX agonists and antagonists may be also be used in drug screens. AAC78449 to
XX AAC78457 and AAB44240 represent sequences used in the exemplification of
XX the present invention.
XX
XX Sequence 602 AA;
XX
XX Query Match 23.5%; Score 63; DB 21; Length 602;
XX Best Local Similarity 42.1%; Pred. No. 1.1;
XX Matches 16; Conservative 7; Mismatches 13; Indels 2; Gaps 2;
XX
XX QY 12 PHTQEAQKQKQSL-GEMPGHLSMSLAL-CLVPLV 47
XX | : : : : | | | | | : | | | | |
XX 559 praedlaedsgsslyrapgrhtwslilaalacvlpll 596
XX
XX RESULT 6
XX AAW56131
XX ID AAW56131 standard; Protein; 1239 AA.
XX
XX AC AAW56131;
XX
XX 17-AUG-1998 (first entry)
XX
XX Drosophila melanogaster KUZ protein.
XX
XX kuzbanian; kuz; neurogenic; KUZ protein; neuronal partitioning;
XX KW development; NOTCH protein processing; regulation; cell function;
XX KW signal transduction pathways; screening; receptor binding;
XX metalloprotease.
XX
XX Drosophila melanogaster.
XX
XX WO9808933-A1.
XX
XX 05-MAR-1998.
XX
XX 27-AUG-1997; 97WO-US15099.
XX
XX
```

```
XX 23-JUL-1997; 97US-0019390.
XX 29-AUG-1996; 96US-0019390.
XX
XX (REGC ) UNIV CALIFORNIA.
XX (UYIA ) UNIV YALE.
XX
XX Pan D, Rooke J, Rubin GM, Xu T, Yavari R;
XX
XX WPI: 1998-179428/16.
XX N-PSDB; AAV22647.
XX
XX New KUZ polypeptides, members of the ADAM family of metalloprotease
XX - useful in neural partitioning and development
XX
XX Claim 1; Pages 31-34; 58pp; English.
XX
XX The sequence is that of encoded by the kuzbanian gene, a
XX neurogenic gene. The KUZ protein is involved in neuronal
XX partitioning and development. It is also involved in processing
XX of the NOTCH protein by cleaving the C-terminal portion. The
XX KUZ protein can be used to screen for compounds that alter
XX binding of KUZ to its receptor or the cleavage of the NOTCH
XX protein, hence acting to regulate NOTCH signal transduction
XX pathways and regulate cell functions.
XX
XX Sequence 1239 AA;
XX
XX Query Match 22.6%; Score 60.5; DB 19; Length 1239;
XX Best Local Similarity 47.2%; Pred. No. 6.2;
XX Matches 17; Conservative 4; Mismatches 10; Indels 5; Gaps 2;
XX
XX QY 9 EKYPTHQEAQKQKQSLGEM----PGHLSMSLA 40
XX | | | | | : | | | : | | | | | : |
XX Db 341 kfyph-qktykeanfaegafydpstgriggssanva 375
XX
XX RESULT 7
XX AAY07092
XX ID AAY07092 standard; Protein; 897 AA.
XX
XX AC AAY07092;
XX
XX 02-JUL-1999 (first entry)
XX
XX Colon cancer associated antigen precursor sequence.
XX
XX Cancer associated antigen; diagnosis; research; treatment; human;
XX KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
XX KW prostate cancer.
XX
XX Homo sapiens.
XX
XX WO9904265-A2.
XX
XX 28-JAN-1999.
XX
XX 15-JUL-1998; 98WO-US14679.
XX
XX 22-JUN-1998; 98US-0102322.
XX 17-JUL-1997; 97US-0896164.
XX 10-OCT-1997; 97US-0061599.
XX 10-OCT-1997; 97US-0061765.
XX 10-OCT-1997; 97US-0948705.
XX 11-OCT-1997; 97GB-0021697.
XX
XX (LUDW-) LUDWIG INST CANCER RES.
XX
XX Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
XX Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
XX Tureci O;
XX
```


05-APR-1999; 9905-0127728.
30-MAR-2000; 2000US-0540763.
(CURA-) CURAGEN CORP.
Shimkets RA, Leach M;
WPI; 2000-602362/57.
N-PSDB; AAC74480.
Novel nucleic acids and peptides derived from open reading frame X,
useful for treating e.g. cancers, proliferative disorders,
neurodegenerative disorders and cardiovascular disease -
Claim 11; Page 455; 5507pp; English.
AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
which represent the human ORFX open reading frames 1 to 3161. The ORFX
sequences have activities such as: cytostatic; hepatotrophic; vulnery;
antipsoratic; antiparkinsonian; nootropic; neuroprotective;
osteopathic; anticoagulant; antiarthritic; immunosuppressant;
immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;
antidiabetic; hypotensive; dermatological; immunosuppressive;
antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
antithyroid; and antianemic. The sequences can be used for determining
the presence of or predisposition to, or preventing or treating
pathological conditions associated with an ORFX-associated disorder. The
nucleic acids can be used to express ORFX proteins in gene therapy
vectors. The proteins and nucleic acids may be used to treat cancers,
proliferative disorders, neurodegenerative disorders, osteoarthritis,
graft vs host disease, cardiovascular disease, diabetes mellitus,
hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
bacterial or fungal infection, malaria, autoimmune disorders, asthma,
allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
nocturnal haemoglobinuria, antiinflammatory disease; to enhance
coagulation; to inhibit thrombosis; and as a contraceptive.

RESULT	10	
AAB40271		
ID	AAB40271	standard; Protein; 184 AA.
XX	XX	
XX	AC	
XX	AAB40271;	
XX	XX	
XX	08-FEB-2001	(first entry)
DT	XX	
DE	Human ORF35	polypeptide sequence SEQ ID NO:70.
XX	XX	
KW	Human; open reading frame; ORF35	detection; cytostatic; hepatotropic;
KW	vulnery; antipsoriatic; antiparkinsonian;	nootropic; neuroprotective;
KW	anticonvulsant; osteopathic; antiarthritic;	immunosuppressant; cardiant;
KW	immunostimulant; thrombolytic; coagulant;	vasotropic; antidiabetic;
KW	hypotensive; dermatological;	immunosuppressive; antinflammatory;
KW	antiviral; antibacterial; antifungal;	antirheumatic; antithyroid;
KW	antianemic; gene therapy; cancer;	proliferative disorder; hypertension;
KW	neurodegenerative disorder; osteoarthritis;	graft vs host disease;
KW	cardiovascular disease; diabetes mellitus;	hypothyroidism; SCID; AIDS;
KW	cholesterol ester storage; systemic lupus	erythematosus; infection;
KW	severe combined immunodeficiency; malaria;	autoimmune disorder; asthma;
KW	allergy; aplastic anaemia; nocturnal	haemoglobinuria; burn; wound;
KW	bone damage; cartilage damage;	antinflammatory disease; coagulation;
KW	thrombosis; contraceptive.	
XX	XX	
OS	Homo sapiens.	
XX	XX	
PN	WO200058473-A2.	
XX	XX	
PD	05-OCT-2000.	
XX	XX	
PF	31-MAR-2000; 2000WO-US08621.	
XX	XX	
PR	31-MAR-1999; 99US-0127607.	
PR	02-APR-1999; 99US-0127636.	

reticul; antiprolact; antipainkiller; noctropic; neuroprotective; anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antitubercular; antifungal; antirheumatic; antihypert; antianaemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive.

RESULT	11
AAY85565	
ID	AAY85565 standard; Protein; 2432 AA.
XX	
XX	AAY85565;
XX	
XX	07-JUL-2000 (first entry)
DT	
XX	
DE	Human homologue of UNC-53 (Hs-UNC-53/2) sequence.
XX	
XX	UNC-53; Caenorhabditis elegans; microtubule; neural regeneration;
KW	anticancer; anti-neurodegeneration; antifibrotic; anti-adhesive; human;
KW	antisclerotic; antimetastatic; anti-arthritis; autoimmune disease.
XX	
XX	Homo sapiens.
OS	
XX	
PH	Location/Qualifiers
Key	
Region	1..89
FT	/note= "this region can be replaced with one of the
FT	three sequences shown in AAY85566 to AAY85568;
FT	this creates three variants at the N-terminus"
FT	
FT	Misc-difference 1018
FT	/label= Asp or Glu
FT	

KW autoimmune disease; hepatic disease; renal disease; allergy; restenosis;
 KW ischaemic shock; Alzheimer's disease; cognitive disorder; schizophrenia;
 KW cardiovascular disorder; wound healing; stroke; arthritis; obesity;
 KW asthma; sepsis; acne; psoriasis; transplant rejection; infection; AIDS;
 KW metabolic disorder.

XX Homo sapiens.

OS

PN WO9938881-A1.

XX

PD 05-AUG-1999.

XX

PF 27-JAN-1999; 99WO-US01621.

XX

PR 30-JAN-1998; 98US-0073170.

XX

PR 30-JAN-1998; 98US-0073159.

PR

PR 30-JAN-1998; 98US-0073160.

PR

PR 30-JAN-1998; 98US-0073161.

PR

PR 30-JAN-1998; 98US-0073162.

PR

PR 30-JAN-1998; 98US-0073164.

PR

PR 30-JAN-1998; 98US-0073165.

PR

PR 30-JAN-1998; 98US-0073167.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Carter KC, Endress GA, Feng P, Ferrie AM, Florence C;

PI Florence KA, Janat F, Ni J, Rosen CA, Ruben SM;

PI Soppet DR, Young P, Yu G;

XX

DR WPI; 1999-469315/39.

DR N-PSDB; AA200412.

XX

PT New isolated human genes and the secreted polypeptides they encode

PT useful in, e.g. treatment of Alzheimer's

XX

PS Claim 1b; Page 300; 393pp; English.

XX

CC This invention describes novel human genes (see AA200410-200477) and the

CC secreted proteins (see AA25711-Y25778) and fragments (see

CC AA25779-Y25907) they encode. The polynucleotides and their corresponding

CC secreted polypeptides are useful for preventing, treating or ameliorating

CC medical conditions e.g. by protein or gene therapy. Also pathological

CC conditions can be diagnosed by determining the amount of the new

CC polypeptides in a sample or by determining the presence of mutations in

CC the new polynucleotides. Specific uses are described for each of the 67

CC polynucleotides of the invention, based on which tissues they are most

CC highly expressed in, and include developing products for the diagnosis

CC or treatment of cancer, tumours, neurodegenerative disorders,

CC developmental abnormalities and fetal deficiencies, blood disorders,

CC leukemias, diseases of the immune system, autoimmune diseases, hepatic

CC and renal disease, inflammation, allergies, ischaemic shock, Alzheimer's

CC and cognitive disorders, schizophrenia, restenosis, cardiovascular

CC disorders, wound healing, stroke, arthritis, obesity, asthma, sepsis,

CC acne, psoriasis, transplant rejection, metabolic disorders, infections

CC and AIDS. The polypeptides are also useful for identifying their binding

CC partners.

XX

SQ Sequence 100 AA;

XX

Query Match

Best Local Similarity 20.3%; Score 54.5; DB 20; Length 100;

Matches 18; Conservative 7; Mismatches 20; Indels 5; Gaps 2;

XX

QY 1 MFLHISPPKYPHTQEAQKRAQSLGEMPGRHLSGSMALCLVPLVREG 50

Db 16 IFlhtrtfvp--vhavkdsaqvleevkheigsqvl---lspveepg 60

XX

RESULT 14

AA98341

ID AAR98341 standard; Protein; 461 AA.

XX

AC AAR98341;

XX 26-NOV-1996 (first entry)

XX Human p57 protein.

XX

XX WD40 repeated structure region; human p57 protein; dimer;

XX anti-cancer; anti-HIV agent; actin; cell movement.

XX Homo sapiens.

XX

XX Key Location/Qualifiers

FT Peptide 77..298

FT /note= "WD40 repeat region, see AAR98339"

FT Peptide 433..461

FT /note= "Leucine rich C-terminal peptide, see also AAR98340"

XX JP08119996-A.

XX

XX 14-MAY-1996.

XX

XX 21-OCT-1994; 94JP-0282743.

XX

XX 21-OCT-1994; 94JP-0282743.

XX

XX (NISB) JAPAN TOBACCO INC.

XX

XX WPI; 1996-283507/29.

XX N-PSDB; AAT30360.

XX

XX A new protein, p57, comprising WD40 repeat region - used for

XX development of anti-cancer and anti-HIV agents

XX

PS Claim 9; Page 23-25; 51pp; Japanese.

XX

CC This sequence represents the human p57 protein. p57 contains a leucine

CC rich C-terminal peptide which comprises a leucine residue after

CC each 7 amino acids (see also R983340) and a WD40 repeated structure

CC region (see also AAR98334-38) containing five WD40 regions. p57 forms a

CC dimer. It can be used in the development of an anti-cancer agent and

CC an anti-HIV agent. It can also combine with actin in the control of

XX cell movement.

XX

SQ Sequence 461 AA;

XX

Query Match 20.3%; Score 54.5; DB 17; Length 461;

Best Local Similarity 43.8%; Pred. No. 14;

Matches 14; Conservative 3; Mismatches 8; Indels 7; Gaps 1;

XX

QY 2 FLHISPPKYPHTQEAQKRAQSLGEMPGRH 33

Db 303 flhlysmf-----sskesqrgmgympkrgl 327

XX

RESULT 15

AA98342

ID AAR98342 standard; Protein; 461 AA.

XX

XX AAR98342;

XX

XX 26-NOV-1996 (first entry)

XX

XX Bovine p57 protein.

XX

XX WD40 repeated structure region; bovine p57 protein; dimer;

XX anti-cancer; anti-HIV agent; actin; cell movement; ds.

XX

XX Bos taurus.

XX

XX Key Location/Qualifiers

FT Peptide 77..108

FT Peptide /note= "WD40 repeat fragment"
FT 127..161
FT /note= "WD40 repeat fragment"
FT 172..203
FT /note= "WD40 repeat fragment"
FT 215..252
FT /note= "WD40 repeat fragment"
FT 265..298
FT /note= "WD40 repeat fragment"
FT 433..461
FT /note= "Leucine rich C terminal fragment"

XX JP08119996-A.
XX
XX
XX 14-MAY-1996.
XX
XX 21-OCT-1994; 94JP-0282743.
XX
XX 21-OCT-1994; 94JP-0282743.
XX
XX (NISB) JAPAN TOBACCO INC.
XX
XX WPI; 1996-283507/29.
XX N-PSDB; AAT30361.
XX
XX A new protein, p57, comprising WD40 repeat region - used for
XX development of anti-cancer and anti-HIV agents
XX
XX Example 2; Page 27-29; 51pp; Japanese.

XX This sequence represents the bovine p57 protein. p57 contains a leucine
XX rich C-terminal peptide which comprises a leucine residue after each
XX 7 amino acids and a WD40 repeated structure region containing five
XX WD40 regions. p57 forms a dimer. It can be used in the development
XX of an anti-cancer agent and an anti-HIV agent. It can also combine
XX with actin in the control of cell movement.

SQ Sequence 461 AA;

Query Match 20.3%; Score 54.5; DB 17; Length 461;
Best Local Similarity 43.8%; Pred. No. 14;
Matches 14; Conservative 3; Mismatches 8; Indels 7; Gaps 1;
Oy 2 FLHISPPKYPHTQEAOKAORSILGEMPGRHLL 33
||| | | : ||: | | | |
Db 303 flhylsmf-----sskesqrgmgypkrgl 327

Search completed: July 30, 2001, 16:51:41
Job time: 216 sec

THIS PAGE BLANK (USPTO)